(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 June 2002 (13.06.2002)

PCT

(10) International Publication Number
WO 02/46385 A2

(51) International Patent Classification[7]: C12N 9/00

(21) International Application Number: PCT/US01/47432

(22) International Filing Date: 4 December 2001 (04.12.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/251,824	7 December 2000 (07.12.2000)	US
60/254,312	8 December 2000 (08.12.2000)	US
60/255,773	14 December 2000 (14.12.2000)	US
60/256,188	15 December 2000 (15.12.2000)	US
60/255,940	15 December 2000 (15.12.2000)	US
60/257,488	21 December 2000 (21.12.2000)	US
60/262,839	19 January 2001 (19.01.2001)	US
60/264,402	26 January 2001 (26.01.2001)	US

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **GRIFFIN, Jennifer, A.** [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **LEE, Ernestine, A.** [US/US]; 624 Kains Street, Albany, CA 94706 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **DUGGAN, Brendan, M.** [AU/US]; 243 Buena Vista Avenue #306, Sunnyvale, CA 94086 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street, #205, San Leandro, CA 94577 (US). **LEE, Sally** [US/US]; 825 East Evelyn, #425, Sunnyvale, CA 94086 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **WARREN, Bridget, A.** [US/US]; 10130 Parkwood Drive #2, Cupertino, CA 95014 (US). **GANDHI, Ameena, R.** [US/US]; 705 5th Avenue, San Francisco, CA 94118 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). **YAO, Monique, G.** [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). **DING, Li** [CN/US]; 3353 Alma Street #146, Palo Alto, CA 94306 (US). **TRIBOULEY, Catherine, M.** [FR/US]; 1121 Tennessee Street, #5, San Francisco, CA 94107 (US). **SANJANWALA, Madhu, M.** [US/US]; 210 Sylvia Court, Los Altos, CA 94024 (US). **ARVIZU, Chandra** [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). **HILLMAN, Jennifer, L.** [US/US]; 230 Monroe Drive, #17, Mountain View, CA 94040 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/46385 A2

(54) Title: ENZYMES

(57) Abstract: The invention provides human enzymes (NZMS) and polynucleotides which identify and encode NZMS. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of NZMS.

ENZYMES

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of enzymes and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative and autoimmune/inflammatory, cardiovascular, gastrointestinal, neurological, pulmonary, reproductive, and eye disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of enzymes.

BACKGROUND OF THE INVENTION

Hydrolases

Hydrolysis is the breaking of a covalent bond in a substrate by introduction of a water molecule. The reaction is catalyzed by a hydrolytic enzyme, or hydrolase, and involves a nucleophilic attack by the water molecule's oxygen atom on a target bond in the substrate. The water molecule is split across the target bond, breaking the bond and generating two product molecules. Hydrolysis reactions form the basis of most metabolic pathways and are present in most biosynthetic pathways. Energy produced in the cell, for example, comes from the hydrolysis of ATP. Hydrolases also participate in reactions essential to functions such as cell signaling, cell proliferation, inflammation, apoptosis, secretion and excretion. Hydrolases are involved in key steps in disease processes involving these functions. Hydrolases may be grouped by substrate specificity into classes including aminohydrolases, phospholipases, carboxyl-esterases, phosphodiesterases, lysozymes, glycosidases, glyoxalases, sulfatases, phosphohydrolases, peptidases, nucleotidases and many others.

Serine hydrolases are a functional class of hydrolytic enzymes that contain a serine residue in their active site. This class of enzymes contains proteinases, esterases, and lipases which hydrolyze a variety of substrates and, therefore, have different biological roles. Proteins in this superfamily can be further grouped into subfamilies based on substrate specificity or amino acid similarities (Puente, X.S. and Lopez-Ont, C. (1995) J. Biol. Chem. 270: 12926-12932). DHH phosphoesterases include the prune protein (Aravind, L. and Koonin, E. V. (1998) Trends Biochem. Sci. 23:17-19).

Carboxylesterases are proteins that hydrolyze carboxylic esters and are classified into three categories- A, B, and C. Most type-B carboxylesterases are evolutionarily related and are considered to comprise a family of proteins. The type-B carboxylesterase family of proteins includes vertebrate acetylcholinesterase, mammalian liver microsomal carboxylesterase, mammalian bile-salt-activated lipase, and duck fatty acyl-CoA hydrolase. Some members of this protein family are not catalytically

active but contain a domain related evolutionarily to other type-B carboxylesterases, such as thyroglobulin and *Drosphila* protein neuractin.

Nucleotidases catalyze the formation of free nucleosides from nucleotides. The cytosolic nucleotidase cN-I (5' nucleotidase-I) cloned from pigeon heart catalyzes the formation of adenosine from AMP generated during ATP hydrolysis (Sala-Newby, G.B. et al. (1999) J. Biol. Chem. 274:17789-17793). Increased adenosine concentration is thought to be a signal of metabolic stress, and adenosine receptors mediate effects including vasodilation, decreased stimulatory neuron firing and ischemic preconditioning in the heart (Schrader, J. (1990) Circulation 81:389-391; Rubino, A. et al. (1992) Eur. J. Pharmacol. 220:95-98; de Jong, J.W. et al. (2000) Pharmacol. Ther. 87:141-149). Deficiency of pyrimidine 5'-nucleotidase can result in hereditary hemolytic anemia (OMIM Entry 266120).

ADP-ribosylation is a reversible post-translational protein modification in which an ADP-ribose moiety is transferred from β-NAD to a target amino acid such as arginine or cysteine. ADP-ribosylarginine hydrolases regenerate arginine by removing ADP-ribose from the protein, completing the ADP-ribosylation cycle (Moss, J. et al. (1997) Adv. Exp. Med. Biol. 419:25-33). ADP-ribosylation is a well-known reaction among bacterial toxins. Cholera toxin, for example, disrupts the adenylyl cyclase system by ADP-ribosylating the α-subunit of the stimulatory G-protein, causing an increase in intracellular cAMP (Moss, J. and Vaughan, M. (eds) (1990) ADP-ribosylating Toxins and G-Proteins: Insights into Signal Transduction, American Society for Microbiology, Washington, D.C.). ADP-ribosylation may also have a regulatory function in eukaryotes, affecting such processes as cytoskeletal assembly (Zhou, H. et al. (1996) Arch. Biochem. Biophys. 334:214-222) and cell proliferation in cytotoxic T-cells (Wang, J. et al. (1996) J. Immunol. 156:2819-2827).

ATPases catalyze the hydrolysis of ATP to ADP in a variety of cellular processes. The ATPases Associated with cellular Activities (AAA) family is characterized by a conserved module of 230 amino acids present in one or two copies in each protein. AAAs function in processes including cell cycle regulation, gene expression in yeast and HIV, vesicle-mediated transport, peroxisome assembly, 26S protease function (Confalonieri, F. and Duguet, M. (1995) Bioessays. 17:639-650). SPAF is a AAA-protein specific to early spermatogenesis and malignant conversion (Liu, Y. et al. (2000) Oncogene 19:1579-1588).

Sulfatases catalyse the hydrolysis of sulfate ester bonds from a variety of substrates, including glycosaminoglycans, sulfolipids, and steroid sulfates. Sulfatase deficiencies are the cause of several human diseases, primarily lysosomal storage disorders. Other disorders associated with sulfatases include metachromatic leukodystrophy, a neurological disorder resulting from a deficiency of

arylsulfatase A, and X-linked recessive chronodysplasia punctata, a disorder of cartilage and bone development due to a deficiency of arylsulfatase E. (See Parenti, G. et al. (1997) Curr. Opin, Genet. Dev. 7:386-391 for review.)

Nucleases comprise both enzymes that hydrolyze DNA (DNase) and RNA (RNase). They serve different purposes in nucleic acid metabolism. Nucleases hydrolyze the phosphodiester bonds between adjacent nucleotides either at internal positions (endonucleases) or at the terminal 3' or 5' nucleotide positions (exonucleases). A DNA exonuclease activity in DNA polymerase, for example, serves to remove improperly paired nucleotides attached to the 3'-OH end of the growing DNA strand by the polymerase and thereby serves a "proofreading" function. DNA endonuclease activity is also involved in the excision step of the DNA repair process.

RNases also serve a variety of functions. For example, RNase P is a ribonucleoprotein enzyme which cleaves the 5' end of pre-tRNAs as part of their maturation process. RNase H digests the RNA strand of an RNA/DNA hybrid. Such hybrids occur in cells invaded by retroviruses, and RNase H is an important enzyme in the retroviral replication cycle. Pancreatic RNase secreted by the pancreas into the intestine hydrolyzes RNA present in ingested foods. RNase activity in serum and cell extracts is elevated in a variety of cancers and infectious diseases (Schein, C.H. (1997) Nat. Biotechnol. 15:529-536). Regulation of RNase activity is being investigated as a means to control tumor angiogenesis, allergic reactions, viral infection and replication, and fungal infections.

Lyases

Lyases are a class of enzymes that catalyze the cleavage of C-C, C-O, C-N, C-S, C-(halide), P-O, or other bonds without hydrolysis or oxidation to form two molecules, at least one of which contains a double bond (Stryer, L. (1995) Biochemistry, W.H. Freeman and Co., New York NY, p.620). Under the International Classification of Enzymes (Webb, E.C. (1992) Enzyme Nomenclature 1992: Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes, Academic Press, San Diego CA), lyases form a distinct class designated by the numeral 4 in the first digit of the enzyme number (i.e., EC 4.x.x.x).

Further classification of lyases reflects the type of bond cleaved as well as the nature of the cleaved group. The group of C-C lyases includes carboxyl-lyases (decarboxylases), aldehyde-lyases (aldolases), oxo-acid-lyases, and other lyases. The C-O lyase group includes hydro-lyases, lyases acting on polysaccharides, and other lyases. The C-N lyase group includes ammonia-lyases, amidine-lyases, amine-lyases (deaminases), and other lyases. Lyases are critical components of cellular biochemistry, with roles in metabolic energy production, including fatty acid metabolism and the

tricarboxylic acid cycle, as well as other diverse enzymatic processes.

One important family of lyases are the carbonic anhydrases (CA), also called carbonate dehydratases, which catalyze the hydration of carbon dioxide in the reaction $H_2O + CO_2 \rightleftharpoons HCO_3^- + H^+$. CA accelerates this reaction by a factor of over 10^6 by virtue of a zinc ion located in a deep cleft about 15Å below the protein's surface and co-ordinated to the imidazole groups of three His residues. Water bound to the zinc ion is rapidly converted to HCO_3^-.

Eight enzymatic and evolutionarily related forms of carbonic anhydrase are currently known to exist in humans: three cytosolic isozymes (CAI, CAII, and CAIII), two membrane-bound forms (CAIV and CAVII), a mitochondrial form (CAV), a secreted salivary form (CAVI) and a yet uncharacterized isozyme (Prosite PDOC00146 Eukaryotic-type carbonic anhydrases signature). Though the isoenzymes CAI, CAII, and bovine CAIII have similar secondary structure and polypeptide-chain fold, CAI has 6 tryptophans, CAII has 7 and CAIII has 8 (Boren, K. et al. (1996) Protein Sci. 5:2479-2484). CAII is the predominant CA isoenzyme in the brain of mammals.

CAs participate in a variety of physiological processes that involve pH regulation, CO_2 and HCO_3^- transport, ion transport, and water and electrolyte balance. For example, CAII contributes to H^+ secretion by gastric parietal cells, by renal tubular cells, and by osteoclasts that secrete H^+ to acidify the bone-resorbing compartment. In addition, CAII promotes HCO_3^- secretion by pancreatic duct cells, cilary body epithelium, choroid plexus, salivary gland acinar cells, and distal colonal epithelium, thus playing a role in the production of pancreatic juice, aqueous humor, cerebrospinal fluid, and saliva, and contributing to electrolyte and water balance. CAII also promotes CO_2 exchange in proximal tubules in the kidney, in erythrocytes, and in lung. CAIV has roles in several tissues: it facilitates HCO_3^- reabsorption in the kidney; promotes CO_2 flux in tissues including brain, skeletal muscle, and heart muscle; and promotes CO_2 exchange from the blood to the alveoli in the lung. CAVI probably plays a role in pH regulation in saliva, along with CAII, and may have a protective effect in the esophagus and stomach. Mitochondrial CAV appears to play important roles in gluconeogenesis and ureagenesis, based on the effects of CA inhibitors on these pathways. (Sly, W.S. and Hu, P.Y. (1995) Ann. Rev. Biochem. 64:375-401.)

A number of disease states are marked by variations in CA activity. Mutations in CAII which lead to CAII deficiency are the cause of osteopetrosis with renal tubular acidosis (Online Medelian Inheritance in Man 259730 Osteopetrosis with Renal Tubular Acidosis). The concentration of CAII in the cerebrospinal fluid (CSF) appears to mark disease activity in patients with brain damage. High CA concentrations have been observed in patients with brain infarction. Patients with transient ischemic attack, multiple sclerosis, or epilepsy usually have CAII concentrations in the normal range,

but higher CAII levels have been observed in the CSF of those with central nervous system infection, dementia, or trigeminal neuralgia (Parkkila, A.K. et al. (1997) Eur. J. Clin. Invest. 27:392-397). Colonic adenomas and adenocarcinomas have been observed to fail to stain for CA, whereas non-neoplastic controls showed CAI and CAII in the cytoplasm of the columnar cells lining the upper half of colonic crypts. The neoplasms show staining patterns similar to less mature cells lining the base of normal crypts (Gramlich T.L. et al. (1990) Arch. Pathol. Lab. Med. 114:415-419).

Therapeutic interventions in a number of diseases involve altering CA activity. CA inhibitors such as acetazolamide are used in the treatment of glaucoma (Stewart, W.C. (1999) Curr. Opin. Opthamol. 10:99-108), essential tremor and Parkinson's disease (Uitti, R.J. (1998) Geriatrics 53:46-48, 53-57), intermittent ataxia (Singhvi, J.P. et al. (2000) Neurology India 48:78-80), and altitude related illnesses (Klocke, D.L. et al. (1998) Mayo Clin. Proc. 73:988-992).

CA activity can be particularly useful as an indicator of long-term disease condition, since the enzyme reacts relatively slowly to physiological changes. CAI and zinc concentrations have been observed to decrease in hyperthyroid Graves' disease (Yoshida, K. (1996) Tohoku J. Exp. Med. 178:345-356) and glycosylated CAI is observed in diabetes mellitus (Kondo, T. et al. (1987) Clin. Chim. Acta 166:227-236). A positive correlation has been observed between CAI and CAII reactivity and endometriosis (Brinton, D.A. et al. (1996) Ann. Clin. Lab. Sci. 26:409-420; D'Cruz , O.J. et al. (1996) Fertil. Steril. 66:547-556).

Another important member of the lyase family is ornithine decarboxylase (ODC), the initial rate-limiting enzyme in polyamine biosynthesis. ODC catalyses the transformation of ornithine into putrescine in the reaction L-ornithine $\rightleftharpoons$ putrescine + CO_2. Polyamines, which include putrescine and the subsequent metabolic pathway products spermidine and spermine, are ubiquitous cell components essential for DNA synthesis, cell differentiation, and proliferation. Thus the polyamines play a key role in tumor proliferation (Medina, M.A. et al. (1999) Biochem. Pharmacol. 57:1341-1344).

ODC is a pyridoxal-5'-phosphate (PLP)-dependent enzyme which is active as a homodimer. Conserved residues include those at the PLP binding site and a stretch of glycine residues thought to be part of a substrate binding region (Prosite PDOC00685 Orn/DAP/Arg decarboxylase family 2 signatures). Mammalian ODCs also contain PEST regions, sequence fragments enriched in proline, glutamic acid, serine, and threonine residues that act as signals for intracellular degradation (Medina, supra).

Many chemical carcinogens and tumor promoters increase ODC levels and activity. Several known oncogenes may increase ODC levels by enhancing transcription of the ODC gene, and ODC itself may act as an oncogene when expressed at very high levels. A high level of ODC is found in a

number of precancerous conditions, and elevation of ODC levels has been used as part of a screen for tumor-promoting compounds (Pegg, A.E. et al. (1995) J. Cell. Biochem. Suppl. 22:132-138).

Inhibitors of ODC have been used to treat tumors in animal models and human clinical trials, and have been shown to reduce development of tumors of the bladder, brain, esophagus, gastrointestinal tract, lung, oral cavity, mammary gland, stomach, skin and trachea (Pegg, supra; McCann, P.P. and Pegg, A.E. (1992) Pharmac. Ther. 54:195-215). ODC also shows promise as a target for chemoprevention (Pegg, supra). ODC inhibitors have also been used to treat infections by African trypanosomes , malaria, and Pneumocystis carinii, and are potentially useful for treatment of autoimmune diseases such as lupus and rheumatoid arthritis (McCann, supra).

Another family of pyridoxal-dependent decarboxylases are the group II decarboxylases. This family includes glutamate decarboxylase (GAD) which catalyzes the decarboxylation of glutamate into the neurotransmitter GABA; histidine decarboxylase (HDC), which catalyzes the decarboxylation of histidine to histamine; aromatic-L-amino-acid decarboxylase (DDC), also known as L-dopa decarboxylase or tryptophan decarboxylase, which catalyzes the decarboxylation of tryptophan to tryptamine and also acts on 5-hydroxy-tryptophan and dihydroxyphenylalanine (L-dopa); and cysteine sulfinic acid decarboxylase (CSD), the rate-limiting enzyme in the synthesis of taurine from cysteine (PROSITE PDOC00329 DDC/GAD/HDC/TyrDC pyridoxal-phosphate attachment site). Taurine is an abundant sulfonic amino acid in brain and is thought to act as an osmoregulator in brain cells (Bitoun, M. and Tappaz, M. (2000) J. Neurochem. 75:919-924).

Phosphatases hydrolytically remove phosphate groups from proteins, an energy-providing step that regulates many cellular processes, including intracellular signaling pathways that in turn control cell growth and differentiation, cell-cell contact, the cell cycle, and oncogenesis.

Peptidases, also called proteases, cleave peptide bonds that form the backbone of peptide or protein chains. Proteolytic processing is essential to cell growth, differentiation, remodeling, and homeostasis as well as inflammation and the immune response. Since typical protein half-lives range from hours to a few days, peptidases are continually cleaving precursor proteins to their active form, removing signal sequences from targeted proteins, and degrading aged or defective proteins. Peptidases function in bacterial, parasitic, and viral invasion and replication within a host. Examples of peptidases include trypsin and chymotrypsin, components of the complement cascade and the blood-clotting cascade, lysosomal cathepsins, calpains, pepsin, renin, and chymosin (Beynon, R.J. and J.S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York, NY, pp. 1-5).

Lysophospholipases (LPLs) regulate intracellular lipids by catalyzing the hydrolysis of ester

bonds to remove an acyl group, a key step in lipid degradation. Small LPL isoforms, approximately 15-30 kD, function as hydrolases; larger isoforms function both as hydrolases and transacylases. A particular substrate for LPLs, lysophosphatidylcholine, causes lysis of cell membranes. LPL activity is regulated by signaling molecules important in numerous pathways, including the inflammatory response.

Thiolester hydrolases, also known as thioesterases, comprise another family of enzymes involved in lipid metabolism. These enzymes have been found in liver, kidney, heart, lung, testis and white and brown adipose tissues, as well as intestine and adrenal gland tissues. Nomenclature of some members of the thioesterase family is derived from demonstration of their compartmentalization within these tissues in the cytosol (CTE), in peroxisomes (PTE) and in mitochondria (MTE) (Hunt, M.C. et al. (1999) J. Biol. Chem. 274:34317-34326). In general, thioesterases participate in the hydrolysis of long chain fatty acids. Acyl-CoA thioesterases catalyze the hydrolysis of acyl-CoA molecules to free fatty acids and CoA. This enzymatic activity is an intrinsic component of animal fatty acid synthetase and in this context serves to terminate chain elongation (Jones, J.M. et al. (1999) J. Biol. Chem. 274:9216-9223). The ability of thioesterases to regulate acyl-CoA concentration in the cell may provide a mechanism for the control of lipid metabolism (Poupon, V. et al. (1999) J. Biol. Chem. 274:19188-19194).

The phosphodiesterases catalyze the hydrolysis of one of the two ester bonds in a phosphodiester compound. Phosphodiesterases are therefore crucial to a variety of cellular processes. Phosphodiesterases include DNA and RNA endo- and exo-nucleases, which are essential to cell growth and replication as well as protein synthesis. Another phosphodiesterase is acid sphingomyelinase, which hydrolyzes the membrane phospholipid sphingomyelin to ceramide and phosphorylcholine. Phosphorylcholine is used in the synthesis of phosphatidylcholine, which is involved in numerous intracellular signaling pathways. Ceramide is an essential precursor for the generation of gangliosides, membrane lipids found in high concentration in neural tissue. Defective acid sphingomyelinase phosphodiesterase leads to a build-up of sphingomyelin molecules in lysosomes, resulting in Niemann-Pick disease.

Glycosidases catalyze the cleavage of hemiacetyl bonds of glycosides, which are compounds that contain one or more sugar. Mammalian lactase-phlorizin hydrolase, for example, is an intestinal enzyme that splits lactose. Mammalian beta-galactosidase removes the terminal galactose from gangliosides, glycoproteins, and glycosaminoglycans, and deficiency of this enzyme is associated with a gangliosidosis known as Morquio disease type B. Vertebrate lysosomal alpha-glucosidase, which hydrolyzes glycogen, maltose, and isomaltose, and vertebrate intestinal sucrase-isomaltase, which

hydrolyzes sucrose, maltose, and isomaltose, are widely distributed members of this family with highly conserved sequences at their active sites.

Phosphoenolpyruvate carboxykinase (ATP) (EC 4.1.1.49) is a lyase involved in gluconeogenesis, the production of glucose from storage compounds in the body. This enzyme catalyzes the decarboxylation of oxaloacetate to form phosphoenolpyruvate, accompanied by hydrolysis of ATP. (See, e.g., Matte, A. et al. (1997) J. Biol. Chem. 272:8105-8108; Medina, V. et al. (1990) J. Bacteriol. 172:7151-7156.)

L-rhamnose and D-fucose are 6-deoxyhexoses found in complex carbohydrates in bacterial cell walls. One of the steps in the pathways leading to the synthesis of these carbohydrates is the conversion of dTDP-D-glucose to an unstable 4-keto-6-deoxy intermediate, a reaction catalyzed by the lyase dTDP-D-glucose 4,6-dehydratase (EC 4.2.1.46). (See, e.g., Tonetti, M. et al. (1998) Biochimie 80:923-931; Yoshida, Y. et al. (1999) J. Biol. Chem. 274:16933-16939.)

Isocitrate lyase (EC 4.1.3.1) is involved in the glyoxylate cycle, a modification of the citric acid cycle. The glyoxylate cycle occurs in bacteria, fungi, and plants. Isocitrate lyase catalyzes the cleavage of isocitrate to yield succinate and glyoxylate. (See, e.g., Beeching, J.R. (1989) Protein Seq. Data Anal. 2:463-466; Atomi, H. et al. (1990) J. Biochem. 107:262-266.)

Aldolases are lyases which catalyze aldol condensation reactions. Fructose 1,6-bisphosphate aldolase (FBP-aldolase; EC 4.1.2.13) catalyzes the reversible cleavage of fructose 1,6-bisphosphate to yield dihydroxyacetone phosphate, a ketose, and glyceraldehyde 3-phosphate, an aldose. Class I FBP-aldolases are found in higher organisms, and exist as homotetramers. Class II FBP-aldolases tend to be dimeric, occur in yeast and bacteria, and have an absolute requirement for a divalent cation for catalytic activity. (See, e.g., Hall, D.R. et al. (1999) J. Mol. Biol. 287:383-394.)

Pseudouridine is an isomer of uridine which helps to maintain the specific tertiary structures of certain rRNAs, tRNAs, and small nuclear and nucleolar RNAs. Pseudouridine is not directly incorporated into these RNAs, but is synthesized by pseudouridine synthases (EC 4.2.1.70), lyases which act on specific uridine residues within these RNAs. The Rlu family of pseudouridine synthases includes Escherichia coli ribosomal large subunit synthase A, which synthesizes pseudouridine at position 746 in 23S rRNA and Escherichia coli ribosomal large subunit synthase C, which synthesizes pseudouridine at positions 955, 2504, and 2580 in 23S rRNA. (See, e.g., Conrad, J. et al. (1998) J. Biol. Chem. 273:18562-18566.)

Fumarate lyases are a group of lyases which share limited sequence homology and use fumarate as a substrate. These enzymes include fumarase (EC 4.2.1.2), aspartase (EC 4.3.1.1), arginosuccinase (EC 4.3.2.2), and adenylosuccinase (EC 4.3.2.2). (See, e.g., Woods, S.A. et al

(1988) Biochim. Biophys. Acta 954:14-26; Woods, S.A. et al. (1988) FEMS Microbiol. Lett. 51:181-186; Zalkin, H. and J.E. Dixon (1992) Prog. Nucleic Acid Res. Mol. Biol. 42:259-287.)

The glyoxylase system is involved in gluconeogenesis, the production of glucose from storage compounds in the body. It consists of glyoxylase I, which catalyzes the formation of S-D-lactoylglutathione from methyglyoxal, a side product of triose-phosphate energy metabolism, and glyoxylase II, which hydrolyzes S-D-lactoylglutathione to D-lactic acid and reduced glutathione. Glyoxylases are involved in hyperglycemia, non-insulin-dependent diabetes mellitus, the detoxification of bacterial toxins, and in the control of cell proliferation and microtubule assembly.

A small subclass of hydrolases acting on ether bonds includes the thioether hydrolases. *S*-adenosyl-L-homocysteine hydrolase, also known as AdoHcyase or SAHH (PROSITE PDOC00603; EC 3.3.1.1), is a thioether hydrolase first described in rat liver extracts as the activity responsible for the reversible hydrolysis of *S*-adenosyl-L-homocysteine (AdoHcy) to adenosine and homocysteine (Sganga, M.W. et al. (1992) PNAS 89:6328-6332). SAHH is a cytosolic enzyme that has been found in all cells that have been tested, with the exception of Escherichia coli and certain related bacteria (Walker, R.D. et al. (1975) Can. J. Biochem. 53:312-319; Shimizu, S. et al. (1988) FEMS Microbiol. Lett. 51:177-180; Shimizu, S. et al. (1984) Eur. J. Biochem. 141:385-392). SAHH activity is dependent on NAD^+ as a cofactor. Deficiency of SAHH is associated with hypermethioninemia (Online Mendelian Inheritance in Man (OMIM) #180960 Hypermethioninemia), a pathologic condition characterized by neonatal cholestasis, failure to thrive, mental and motor retardation, facial dysmorphism with abnormal hair and teeth, and myocaridopathy (Labrune, P. et al. (1990) J. Pediat. 117:220-226).

Another subclass of hydrolases includes those enzymes which act on carbon-nitrogen (C-N) bonds other than peptide bonds. To this subclass belong those enzymes hydrolyzing amides, amidines, and other C-N bonds. This subclass is further subdivided on the basis of substrate specificity such as linear amides, cyclic amides, linear amidines, cyclic amidines, nitriles and other compounds. A hydrolase belonging to the sub-subclass of enzymes acting only on asparagine-oligosaccharides containing one amino acid is N^4-(β-*N*-acetylglucosaminyl)-L-asparaginase, or aspartylglucosylaminidase (AGA; EC 3.5.1.26). AGA is a key enzyme in the catabolism of N-linked oligosaccharides of glycoproteins. It cleaves the asparagine from the residual N-acetylglucosamines as one of the final steps in the lysosomal breakdown of glycoproteins. AGA is an enzyme of lysosomal origin that has been found in worms, rats, mice, pigs, humans, and flavobacteria (ExPASy Enzyme View of ENZYME: 3.5.1.2; SWISS-PROT P20933). A deficiency of AGA causes a lysosomal disease known as aspartylglucosaminuria (AGU) (Online Mendelian Inheritance in Man

(OMIM) #208400 Aspartylglucosaminuria; Jenner, F.A. et al. (1967) Biochem. J. 103:48P-49P; Pollitt, R.J. et al. (1968) Lancet II:253-255). Patients with AGU exhibit severe mental retardation, cranial asymmetry, scoliosis, periodic hyperactivity, and vacuolated lymphocytes. AGU in infants is characterized by diarrhea and frequent infections (Palo, J. et al. (1970) J. Ment. Defic. Res. 14:168-173). It has been shown that AGU stems from genetic mutations in the AGU gene, which probably affects the folding and stability of the AGA molecule (Ikonen, E. et al. (1991) PNAS 88:11222-11226; Ikonen, E. et al. (1991) EMBO J. 10:51-58; Ikonen, E. et al. (1991) Genomics 11:206-211). Metabolic consequences of AGA deficiency in mice have been found to be associated with defects in neuromotor coordination, including impaired bladder function and severe ataxic gait in older mice (Tenhunen, K. et al. (1995) Genomics 30:244-250; Gonzalez-Gomez, I. et al. (1998) Am. J. Path. 153:1293-1300).

Pancreatic ribonucleases (RNase) are pyrimidine-specific endonucleases found in high quantity in the pancreas of certain mammalian taxa and of some reptiles (Beintema, J.J. et al (1988) Prog. Biophys. Mol. Biol. 51:165-192). Proteins in the mammalian pancreatic RNase superfamily are noncytosolic endonucleases that degrade RNA through a two-step transphosphorolytic-hydrolytic reaction (Beintema, J.J. et al. (1986) Mol. Biol. Evol. 3:262-275). Specifically, the enzymes are involved in endonucleolytic cleavage of 3'-phosphomononucleotides and 3'-phosphooligonucleotides ending in C-P or U-P with 2',3'-cyclic phosphate intermediates. Ribonucleases can unwind the DNA helix by complexing with single-stranded DNA; the complex arises by an extended multi-site cation-anion interaction between lysine and arginine residues of the enzyme and phosphate groups of the nucleotides. Some of the enzymes belonging to this family appear to play a purely digestive role, whereas others exhibit potent and unusual biological activities (D'Alessio, G. (1993) Trends Cell Biol. 3:106-109). Proteins belonging to the pancreatic RNase family include: bovine seminal vesicle and brain ribonucleases; kidney non-secretory ribonucleases (Beintema, J.J. et al (1986) FEBS Lett. 194:338-343); liver-type ribonucleases (Rosenberg, H.F. et al. (1989) PNAS U.S.A. 86:4460-4464); angiogenin, which induces vascularisation of normal and malignant tissues; eosinophil cationic protein (Hofsteenge, J. et al. (1989) Biochemistry 28:9806-9813), a cytotoxin and helminthotoxin with ribonuclease activity; and frog liver ribonuclease and frog sialic acid-binding lectin. The sequences of pancreatic RNases contain 4 conserved disulphide bonds and 3 amino acid residues involved in the catalytic activity.

Aconitase (EC 4.2.1.3) is a lyase which carries out a crucial step in the tricarboxylic acid cycle. Aconitase catalyzes the reversible transformation of citrate into isocitrate through a cis-aconitate intermediate. Two forms of aconitase are found in mammalian cells, a cytosolic aconitase

(Kennedy, M.C. et al. (1992) Proc. Natl. Acad. Sci. USA 89:11730-11734) and a mitochondrial aconitase (Mirel, D.B. et al. (1998) Gene 213:205-218).

Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco; EC 4.1.1.39) is a lyase which carries out a crucial step in the Calvin cycle during photosynthesis. Rubisco catalyzes the covalent incorporation of carbon dioxide into the 5-carbon sugar ribulose 1,5-bisphosphate along with the simultaneous cleavage of this molecule into two molecules of 3-phosphoglycerate. (See, e.g., Hartman, F.C. and M.R. Harpel (1994) Annu. Rev. Biochem. 63:197-234.) Specific methyltransferases (EC 2.1.1.43) catalyze the methylation of amino groups near the N-termini of the small and large subunits of Rubisco (Ying, Z. et al. (1998) Acta Biol. Hung. 49:173-184; Klein, R.R. and R.L. Houtz (1995) Plant Mol. Biol. 27:249-261).

Dihydrodipicolinate synthetase (EC 4.2.1.52) is a lyase involved in lysine biosynthesis. This enzyme catalyzes the condensation of pyruvate and aspartic-β-semialdehyde with the elimination of water to produce 2,3-dihydrodipicolinate.

Proper regulation of lyases is critical to normal physiology. For example, mutation induced deficiencies in the uroporphyrinogen decarboxylase can lead to photosensitive cutaneous lesions in the genetically-linked disorder familial porphyria cutanea tarda (Mendez, M. et al. (1998) Am. J. Genet. 63:1363-1375). It has also been shown that adenosine deaminase (ADA) deficiency stems from genetic mutations in the ADA gene, resulting in the disorder severe combined immunodeficiency disease (SCID) (Hershfield, M.S. (1998) Semin. Hematol. 35:291-298).

The discovery of new enzymes, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative and autoimmune/inflammatory, cardiovascular, gastrointestinal, neurological, pulmonary, reproductive, and eye disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of enzymes.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, enzymes, referred to collectively as "NZMS" and individually as "NZMS-1," "NZMS-2," "NZMS-3," "NZMS-4," "NZMS-5," "NZMS-6," "NZMS-7," "NZMS-8," "NZMS-9," "NZMS-10," "NZMS-11," "NZMS-12," "NZMS-13," "NZMS-14," "NZMS-15," "NZMS-16," "NZMS-17," and "NZMS-18." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the

group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-18.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-18. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:19-36.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a

polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90%

identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional NZMS, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional NZMS, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide

comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional NZMS, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptides of the invention. The probability scores for the matches between

each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"NZMS" refers to the amino acid sequences of substantially purified NZMS obtained from

any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of NZMS. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of NZMS either by directly interacting with NZMS or by acting on components of the biological pathway in which NZMS participates.

An "allelic variant" is an alternative form of the gene encoding NZMS. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding NZMS include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as NZMS or a polypeptide with at least one functional characteristic of NZMS. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding NZMS, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding NZMS. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent NZMS. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of NZMS is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence

to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of NZMS. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of NZMS either by directly interacting with NZMS or by acting on components of the biological pathway in which NZMS participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind NZMS polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an in vitro evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or $2'-NH_2$), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system.

Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

The term "intramer" refers to an aptamer which is expressed in vivo. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic NZMS, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding NZMS or fragments of NZMS may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be

associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of NZMS or the polynucleotide encoding NZMS which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:19-36 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:19-36, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:19-36 is useful, for

example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:19-36 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:19-36 and the region of SEQ ID NO:19-36 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-18 is encoded by a fragment of SEQ ID NO:19-36. A fragment of SEQ ID NO:1-18 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-18. For example, a fragment of SEQ ID NO:1-18 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-18. The precise length of a fragment of SEQ ID NO:1-18 and the region of SEQ ID NO:1-18 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from

several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions,

explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a

complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one

nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of NZMS which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of NZMS which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of NZMS. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of NZMS.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and

may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an NZMS may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of NZMS.

"Probe" refers to nucleic acid sequences encoding NZMS, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas

South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent,

chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing NZMS, nucleic acids encoding NZMS, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based

on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The

presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human enzymes (NZMS), the polynucleotides encoding NZMS, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative and autoimmune/inflammatory, cardiovascular, gastrointestinal, neurological, pulmonary, reproductive, and eye disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog and the PROTEOME database identification numbers (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and

2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are enzymes.

For example, SEQ ID NO:1 is 59% identical to human carbonic anhydrase I (GenBank ID g179793) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 6.5e-87, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains a eukaryotic-type carbonic anhydrase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:1 is a carbonic anhydrase.

As another example, SEQ ID NO:3 is 34% identical to Halobacterium dihydrodipicolinate synthase (GenBank ID g10580053) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 6.0e-29, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:3 also contains a dihydrodipicolinate synthetase family domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:3 is a dihydrodipicolinate synthase.

As another example, SEQ ID NO:4 is 98% identical to Rattus norvegicus S-adenosyl-L-homocysteine hydrolase (GenBank ID g1185363) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.1e-230, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:4 also contains an S-adenosyl-L-homocysteine hydrolase signature pattern as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from additional BLAST and MOTIFS

analyses provide further corroborative evidence that SEQ ID NO:4 is an S-adenosyl-L-homocysteine hydrolase.

As another example, SEQ ID NO:7 is 59% identical to Sanguinus oedipus ribonuclease k6 precursor (GenBank ID g2745760) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 5.3e-44, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:7 also contains a pancreatic ribonuclease domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:7 is a pancreatic ribonuclease.

As another example, SEQ ID NO:10 is 55% identical to human arylsulfatase B precursor (GenBank ID g179077) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.9e-144, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:10 also contains a sulfatase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:10 is a sulfatase.

As another example, SEQ ID NO:13 is 55% identical to feline arylsulfatase B (ARSB) (GenBank ID g258856) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 3.1e-144, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:13 also contains a sulfatase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:13 is a sulfatase.

As another example, SEQ ID NO:14 is 80% identical from residues N200 to K362 to human S-adenosylhomocysteine hydrolase (GenBank ID g178279) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 5.6e-83, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:14 also contains an S-adenosyl-L-homocysteine hydrolase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and additional

BLAST analyses provide further corroborative evidence that SEQ ID NO:14 is an S-adenosyl-L-homocysteine hydrolase. The algorithms and parameters for the analysis of SEQ ID NO:1 are described in Table 7.

As another example, SEQ ID NO:15 is 56% identical to Mus musculus spermatogenesis associated ATPase (GenBank ID g4105619) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 8.6e-144, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:15 also contains an AAA ATPase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:15 is an AAA ATPase.

SEQ ID NO:2, SEQ ID NO:5-6, SEQ ID NO:8-9, SEQ ID NO:11-12 and SEQ ID NO:16-18 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-18 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5′) and stop (3′) positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide sequences of the invention, and of fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:19-36 or that distinguish between SEQ ID NO:19-36 and related polynucleotide sequences.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those

sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_*XXXXXX*_N_1_N_2_*YYYYY*_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	**Type of analysis and/or examples of programs**
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 2 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses NZMS variants. A preferred NZMS variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the NZMS amino acid sequence, and which contains at least one functional or structural characteristic of NZMS.

The invention also encompasses polynucleotides which encode NZMS. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:19-36, which encodes NZMS. The polynucleotide sequences of SEQ ID NO:19-36, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding NZMS. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding NZMS. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:19-36 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:19-36. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of NZMS.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide sequence encoding NZMS. A splice variant may have portions which have significant sequence identity to the polynucleotide sequence encoding NZMS, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to the polynucleotide sequence encoding NZMS over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide

sequence encoding NZMS. For example, a polynucleotide comprising a sequence of SEQ ID NO:35 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:36. Any one of the splice variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of NZMS.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding NZMS, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring NZMS, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode NZMS and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring NZMS under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding NZMS or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding NZMS and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode NZMS and NZMS derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding NZMS or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:19-36 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of

the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding NZMS may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about

68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode NZMS may be cloned in recombinant DNA molecules that direct expression of NZMS, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express NZMS.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter NZMS-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of NZMS, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is

produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding NZMS may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, NZMS itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of NZMS, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active NZMS, the nucleotide sequences encoding NZMS or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding NZMS. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding NZMS. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding NZMS and its initiation codon and upstream regulatory sequences are inserted

into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding NZMS and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding NZMS. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding NZMS. For example, routine cloning,

subcloning, and propagation of polynucleotide sequences encoding NZMS can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding NZMS into the vector's multiple cloning site disrupts the *lac*Z gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of NZMS are needed, e.g. for the production of antibodies, vectors which direct high level expression of NZMS may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of NZMS. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of NZMS. Transcription of sequences encoding NZMS may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding NZMS may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses NZMS in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-

based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of NZMS in cell lines is preferred. For example, sequences encoding NZMS can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ and *apr*⁻ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding NZMS is inserted within a marker gene sequence, transformed cells containing

sequences encoding NZMS can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding NZMS under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding NZMS and that express NZMS may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of NZMS using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on NZMS is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding NZMS include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding NZMS, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding NZMS may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein

produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode NZMS may be designed to contain signal sequences which direct secretion of NZMS through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding NZMS may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric NZMS protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of NZMS activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the NZMS encoding sequence and the heterologous protein sequence, so that NZMS may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled NZMS may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the

T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

NZMS of the present invention or fragments thereof may be used to screen for compounds that specifically bind to NZMS. At least one and up to a plurality of test compounds may be screened for specific binding to NZMS. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of NZMS, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which NZMS binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express NZMS, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing NZMS or cell membrane fractions which contain NZMS are then contacted with a test compound and binding, stimulation, or inhibition of activity of either NZMS or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with NZMS, either in solution or affixed to a solid support, and detecting the binding of NZMS to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

NZMS of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of NZMS. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for NZMS activity, wherein NZMS is combined with at least one test compound, and the activity of NZMS in the presence of a test compound is compared with the activity of NZMS in the absence of the test compound. A change in the activity of NZMS in the presence of the test compound is indicative of a compound that modulates the activity of NZMS. Alternatively, a test compound is combined with an in vitro or cell-free system comprising NZMS under conditions suitable for NZMS activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of NZMS

may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding NZMS or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding NZMS may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding NZMS can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding NZMS is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress NZMS, e.g., by secreting NZMS in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of NZMS and enzymes. In addition, the expression of NZMS is closely associated with brain

tissue, breast tissue, bronchial smooth muscle tissue, endometrial tissue, kidney tissue, liver tissue, lung tissue, pituitary tissue, prostate tissue, small intestine tissue, THP-1 promonocyte cells, and thymus tissue. Therefore, NZMS appears to play a role in cell proliferative and autoimmune/inflammatory, cardiovascular, gastrointestinal, neurological, pulmonary, reproductive, and eye disorders. In the treatment of disorders associated with increased NZMS expression or activity, it is desirable to decrease the expression or activity of NZMS. In the treatment of disorders associated with decreased NZMS expression or activity, it is desirable to increase the expression or activity of NZMS.

Therefore, in one embodiment, NZMS or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NZMS. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary

angioneurotic edema, and immunodeficiency associated with Cushing's disease; a cardiovascular disorder such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery , congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, $alpha_1$-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia,

eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a pulmonary disorder, such as congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian

tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; and an eye disorder such as ocular hypertension and glaucoma.

In another embodiment, a vector capable of expressing NZMS or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NZMS including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified NZMS in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NZMS including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of NZMS may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NZMS including, but not limited to, those listed above.

In a further embodiment, an antagonist of NZMS may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of NZMS. Examples of such disorders include, but are not limited to, those cell proliferative and autoimmune/inflammatory, cardiovascular, gastrointestinal, neurological, pulmonary, reproductive, and eye disorders described above. In one aspect, an antibody which specifically binds NZMS may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express NZMS.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding NZMS may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of NZMS including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of NZMS may be produced using methods which are generally known in the

art. In particular, purified NZMS may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind NZMS. Antibodies to NZMS may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with NZMS or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to NZMS have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of NZMS amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to NZMS may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce NZMS-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton,

D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for NZMS may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between NZMS and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering NZMS epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for NZMS. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of NZMS-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple NZMS epitopes, represents the average affinity, or avidity, of the antibodies for NZMS. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular NZMS epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the NZMS-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of NZMS, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of NZMS-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding NZMS, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding NZMS. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding NZMS. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding NZMS may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene

Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in NZMS expression or regulation causes disease, the expression of NZMS from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in NZMS are treated by constructing mammalian expression vectors encoding NZMS and introducing these vectors by mechanical means into NZMS-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of NZMS include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). NZMS may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding NZMS from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to NZMS expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding NZMS under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding NZMS to cells which have one or more genetic abnormalities with respect to the expression of NZMS. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are

described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding NZMS to target cells which have one or more genetic abnormalities with respect to the expression of NZMS. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing NZMS to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding NZMS to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for NZMS into the alphavirus genome in place of the capsid-coding region results in the production of a large number of NZMS-coding RNAs and the synthesis of high levels of NZMS in vector transduced cells. While

alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of NZMS into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding NZMS.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis.

Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding NZMS. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding NZMS. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased NZMS expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding NZMS may be therapeutically useful, and in the treatment of disorders associated with decreased NZMS expression or activity, a compound which specifically promotes expression of the polynucleotide encoding NZMS may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding NZMS is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted

biochemical system. Alterations in the expression of a polynucleotide encoding NZMS are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding NZMS. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of NZMS, antibodies to NZMS, and mimetics, agonists, antagonists, or inhibitors of NZMS.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical,

sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising NZMS or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, NZMS or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example NZMS or fragments thereof, antibodies of NZMS, and agonists, antagonists or inhibitors of NZMS, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is

preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind NZMS may be used for the diagnosis of disorders characterized by expression of NZMS, or in assays to monitor patients being treated with NZMS or agonists, antagonists, or inhibitors of NZMS. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for NZMS include methods which utilize the antibody and a label to detect NZMS in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring NZMS, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of NZMS expression. Normal or standard values for NZMS expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to NZMS under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of NZMS expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding NZMS may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of NZMS may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of NZMS, and to monitor regulation of NZMS levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding NZMS or closely related molecules may be used to identify nucleic acid sequences which encode NZMS. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding NZMS, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the NZMS encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:19-36 or from genomic sequences including promoters, enhancers, and introns of the NZMS gene.

Means for producing specific hybridization probes for DNAs encoding NZMS include the cloning of polynucleotide sequences encoding NZMS or NZMS derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding NZMS may be used for the diagnosis of disorders associated with expression of NZMS. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory

disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and an eye disorder such as ocular hypertension and glaucoma. The polynucleotide sequences encoding NZMS may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or

tissues from patients to detect altered NZMS expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding NZMS may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding NZMS may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding NZMS in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of NZMS, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding NZMS, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding NZMS

may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding NZMS, or a fragment of a polynucleotide complementary to the polynucleotide encoding NZMS, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding NZMS may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding NZMS are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of NZMS include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray

can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, NZMS, fragments of NZMS, or antibodies specific for NZMS may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000)

Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot

is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for NZMS to quantify the levels of NZMS expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological

sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding NZMS may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding NZMS on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as

linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, NZMS, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between NZMS and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with NZMS, or fragments thereof, and washed. Bound NZMS is then detected by methods well known in the art. Purified NZMS can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding NZMS specifically compete with a test compound for binding NZMS. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with NZMS.

In additional embodiments, the nucleotide sequences which encode NZMS may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding

description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/251,824, U.S. Ser. No. 60/255,773, U.S. Ser. No. 60/254,312, U.S. Ser. No. 60/256,188, U.S. Ser. No. 60/257,488, U.S. Ser. No. 60/262,839, U.S. Ser. No. 60/264,402, U.S. Ser. No. 60/255,940, and U.S. Ser. No. [Attorney Docket No. PF-1245 P, filed October 19, 2001] are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 2. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs

were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading

frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from Homo sapiens, Rattus norvegicus, Mus musculus, Caenorhabditis elegans, Saccharomyces cerevisiae, Schizosaccharomyces pombe, and Candida albicans (Incyte Genomics, Palo Alto CA); and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold

parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:19-36. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative enzymes were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode enzymes, the encoded polypeptides were analyzed by querying against PFAM models for enzymes. Potential enzymes were also identified by homology to Incyte cDNA sequences that had been annotated as enzymes. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of NZMS Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:19-36 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:19-36 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

In this manner, SEQ ID NO:31 was mapped to chromosome 8 within the interval from 64.60 to 78.80 centiMorgans.

In this manner, SEQ ID NO:32 was mapped to chromosome 11 within the interval from 92.50 to 96.20 centiMorgans.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar.

The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{5 \times \text{minimum}\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding NZMS are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding NZMS. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of NZMS Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%)

agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:19-36 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature

under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/µl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/µl RNase inhibitor, 500 µM dATP, 500 µM dGTP, 500 µM dTTP, 40 µM dCTP, 40 µM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with

GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37° C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85° C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 µl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 µg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 µl of the array element DNA, at an average concentration of 100 ng/µl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60° C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample

mixture is heated to 65° C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high

signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI. Complementary Polynucleotides

Sequences complementary to the NZMS-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring NZMS. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of NZMS. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the NZMS-encoding transcript.

XII. Expression of NZMS

Expression and purification of NZMS is achieved using bacterial or virus-based expression systems. For expression of NZMS in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express NZMS upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of NZMS in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding NZMS by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther.

7:1937-1945.)

In most expression systems, NZMS is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from NZMS at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified NZMS obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, and XVIII, etc. where applicable.

XIII. Functional Assays

NZMS function is assessed by expressing the sequences encoding NZMS at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of

fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of NZMS on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding NZMS and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding NZMS and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of NZMS Specific Antibodies

NZMS substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the NZMS amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-NZMS activity by, for example, binding the peptide or NZMS to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XV. Purification of Naturally Occurring NZMS Using Specific Antibodies

Naturally occurring or recombinant NZMS is substantially purified by immunoaffinity chromatography using antibodies specific for NZMS. An immunoaffinity column is constructed by covalently coupling anti-NZMS antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing NZMS are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of NZMS (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/NZMS binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and NZMS is collected.

XVI. Identification of Molecules Which Interact with NZMS

NZMS, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled NZMS, washed, and any wells with labeled NZMS complex are assayed. Data obtained using different concentrations of NZMS are used to calculate values for the number, affinity, and association of NZMS with the candidate molecules.

Alternatively, molecules interacting with NZMS are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

NZMS may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVII. Demonstration of NZMS activity

Lyase activity of NZMS is demonstrated through a variety of specific enzyme assays. In general, NZMS is incubated with its substrate(s) under conditions suitable for the enzymatic reaction being assayed. After a suitable period of time, the reaction is terminated, and the formation of the product(s) are monitored spectrophotometrically, chromatographically, fluorometrically, or by some other appropriate method. Lyase activity is proportional to the amount of product(s) formed, or the rate of product formation. Some examples of specific lyase activity assays are described below.

Glyoxalase activity of NZMS is measured spectrophotometrically as described (Ridderstrom, M. et al. (1996) J. Biol. Chem. 271:319-323). NZMS is added to a 1 ml reaction volume containing 900 μM *S*-D-lactoylglutathione and 200 μM 5,5'-dithiobis(2-nitrobenzoate) in 100 mM MOPS, pH 7.2, at 37 C. The formation of glutathione is monitored spectrophotometrically at 412 nm.

Glyoxalase I activity of NZMS is measured by monitoring the formation of glutathione thioester from methylglyoxal and glutathione. NZMS is incubated with 2mM methylglyoxal and 2 mM reduced glutathione in 0.1 M sodium phosphate, pH 7.0, at 30°C. Formation of the glutathione

thioester is monitored spectrophotometrically at a wavelength of 240 nm. Glyoxalase I activity of NZMS is proportional to the rate of formation of the glutathione thioester. (See, e.g., Ridderstrom, M. et al. (1998) J. Biol. Chem. 273:21623-21628.)

dTDP-D-glucose 4,6-dehydratase activity of NZMS is measured by monitoring the formation of dTDP-4-keto-6-deoxy-D-glucose from dTDP-D-glucose. NZMS is incubated with 50 mM Tris-HCl, pH 7.6, 12 mM $MgCl_2$, 4 mM dTDP-D-glucose, 0.9 unit of inorganic pyrophosphatase, and 8 mM NADPH for 3 hours at 37°C. The sugar components in the mixture are coupled with 2-aminopyridine and then analyzed chromatographically using an anion-exchange column. Dehydratase activity is proportional to the amount of dTDP-4-keto-6-deoxy-D-glucose formed. (See, e.g., Yoshida, 1999, supra.)

Aconitase activity of NZMS is measured in an assay coupled to isocitric dehydrogenase. NZMS is incubated with isocitric dehydrogenase, NADP, and citrate, and the reduction of NADP is monitored fluorometrically. Aconitase activity is proportional to the rate of NADP reduction. (See, e.g., Costello, L.C. et al. (1997) J. Biol. Chem. 272:28875-28881; Costello, L.C. et al. (1996) Urology 48:654-659.)

Dihydrodipicolinate synthase activity of NZMS is measured using the o-aminobenzaldehyde method (Yugari, Y. and C. Gilvarg (1965) J. Biol. Chem. 240:4710-4716; Karchi, H. et al. (1994) Proc. Natl. Acad. Sci. USA 91:2577-2581). Alternatively, dihydrodipicolinate synthase activity of NZMS is measured as described by Frisch and coworkers (Frisch, D.A. et al. (1991) Plant Physiol. 96:444-452; Shaver, J.M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:1962-1966).

Sulfatase activity of NZMS is measured by incubating NZMS with the synthetic substrate *p*-nitrocatechol sulfate and monitoring the release of free *p*-nitrocatechol after the addition of base. The activity of NZMS is proportional to the amount of free *p*-nitrocatechol released, as measured spectrophotometrically at 515 nm.

Ribonuclease activity of NZMS can be measured spectrophotometrically by determining the amount of solubilized RNA that is produced as a result of incubation of RNA substrate with NZMS. 5 µl (20 µg) of a 4 mg/ml solution of yeast tRNA (Sigma) is added to 0.8 ml of 40 mM sodium phosphate, pH 7.5, containing NZMS. The reaction is incubated at 25°C for 15 minutes. The reaction is stopped by addition of 0.5 ml of an ice-cold fresh solution of 20 mM lanthanum nitrate plus 3% perchloric acid. The stopped reaction is incubated on ice for at least 15 min, and the insoluble tRNA is removed by centrifugation for 5 min at 10,000 g. Solubilized tRNA is determined as UV absorbance (260 nm) of the remaining supernatant, with A_{260} of 1.0 corresponding to 40 µg of solubilized RNA (Rosenberg, H.F. et al. (1996) Nucleic Acids Research 24:3507-3513).

An assay for carbonic anhydrase activity of NZMS uses the fluorescent pH indicator 8-hydroxypyrene-1,3,6-trisulfonate (pyranine) in combination with stopped-flow fluorometry to measure carbonic anhydrase activity (Shingles, et al. 1997, Anal. Biochem. 252: 190-197). A pH 6.0 solution is mixed with a pH 8.0 solution and the initial rate of bicarbonate dehydration is measured. Addition of carbonic anhydrase to the pH 6.0 solution enables the measurement of the initial rate of activity at physiological temperatures with resolution times of 2 ms. Shingles et al. used this assay to resolve differences in activity and sensitivity to sulfonamides by comparing mammalian carbonic anhydrase isoforms. The fluorescent technique's sensitivity allows the determination of initial rates with a protein concentration as little as 65 ng/ml.

Decarboxylase activity of NZMS is measured as the release of CO_2 from labeled substrate. For example, ornithine decarboxylase activity of NZMS is assayed by measuring the release of CO_2 from L-[1-^{14}C]-ornithine (Reddy, S.G et al. (1996) J. Biol. Chem. 271:24945-24953). Activity is measured in 200 µl assay buffer (50 mM Tris/HCl, pH 7.5, 0.1 mM EDTA, 2 mM dithiothreitol, 5 mM NaF, 0.1% Brij35, 1 mM PMSF, 60 µM pyridoxal-5-phosphate) containing 0.5 mM L-ornithine plus 0.5 µCi L-[1-^{14}C]ornithine. The reactions are stopped after 15-30 minutes by addition of 1 M citric acid, and the $^{14}CO_2$ evolved is trapped on a paper disk filter saturated with 20 µl of 2 N NaOH. The radioactivity on the disks is determined by liquid scintillation spectography. The amount of $^{14}CO_2$ released is proportional to ornithine decarboxylase activity of NZMS.

Protein phosphatase activity can be measured by the hydrolysis of p-nitrophenyl phosphate (PNPP). NZMS is incubated together with PNPP in HEPES buffer pH 7.5, in the presence of 0.1% β-mercaptoethanol at 37°C for 60 min. The reaction is stopped by the addition of 6 ml of 10 N NaOH (Diamond, R.H. et al. (1994) Mol. Cell. Biol. 14:3752-62). Alternatively, acid phosphatase activity of NZMS is demonstrated by incubating NZMS containing extract with 100 µl of 10 mM PNPP in 0.1 M sodium citrate, pH 4.5, and 50 µl of 40 mM NaCl at 37°C for 20 min. The reaction is stopped by the addition of 0.5 ml of 0.4 M glycine/NaOH, pH 10.4 (Saftig, P. et al. (1997) J. Biol. Chem. 272:18628-18635). The increase in light absorbance at 410 nm resulting from the hydrolysis of PNPP is measured using a spectrophotometer. The increase in light absorbance is proportional to the activity of NZMS in the assay.

In the alternative, NZMS activity is determined by measuring the amount of phosphate removed from a phosphorylated protein substrate. Reactions are performed with 2 or 4 nM NZMS in a final volume of 30 µl containing 60 mM Tris, pH 7.6, 1 mM EDTA, 1 mM EGTA, 0.1% 2-mercaptoethanol and 10 µM substrate, ^{32}P-labeled on serine/threonine or tyrosine, as appropriate. Reactions are initiated with substrate and incubated at 30° C for 10-15 min. Reactions are quenched

with 450 µl of 4% (w/v) activated charcoal in 0.6 M HCl, 90 mM $Na_4P_2O_7$, and 2 mM NaH_2PO_4, then centrifuged at 12,000 × g for 5 min. Acid-soluble ^{32}Pi is quantified by liquid scintillation counting (Sinclair, C. et al. (1999) J. Biol. Chem. 274:23666-23672).

Additionally, NZMS activity can be determined by measuring the amount of sulfate removed from a sulfonated protein substrate. Reactions are performed in 50 mM Tris-HCl buffer, pH 8.0 containing 5 mM 4-nitrocatechol sulfate and 5 µl crude supernatant protein extracted from cells expressing NZMS. The reaction is incubated at 37°C for 30 minutes (Hallmann, A. et al. (1994) Eur. J. Biochem. 221:143-150.) The increase in light absorbance at 410 nm resulting from the hydrolysis of the phenol sulfate substrate is measured using a spectrophotometer. The increase in light absorbance is proportional to the activity of NZMS in the assay.

NZMS activity can be measured by determining the amount of free adenosine produced by the hydrolysis of AMP, as described by Sala-Newby et al. supra. Briefly, NZMS is incubated with AMP in a suitable buffer for 10 minutes at 37°C. Free adenosine is separated from AMP and measured by reverse phase HPLC.

Alternatively, NZMS activity is measured by the NZMSolysis of ADP-ribosylarginine (Konczalik, P. and J. Moss (1999) J. Biol. Chem. 274:16736-16740). 50 ng of NZMS are incubated with 100 µM ADP-ribosyl-[^{14}C]arginine (78,000 cpm) in 50 mM potassium phosphate, pH 7.5, 5 mM dithiothreitol, 10 mM $MgCl_2$ in a final volume of 100 µl. After 1 h at 37° C, 90 µl of the sample is applied to a column (0.5 × 4 cm) of Affi-Gel 601 (boronate) equilibrated and eluted with five 1-ml portions of 0.1 M glycine, pH 9.0, 0.1 M NaCl, and 10 mM $MgCl_2$. Free ^{14}C-Arg in the total eluate is measured by liquid scintillation counting.

NZMS hydrolytic activity is measured in the hydrolytic direction spectroscopically by measuring the rate of the product (homocysteine) formed by reaction with 5,5'-Dithiobis(2-nitrobenzoic acid) (DTNB). To 800 µl of an enzyme solution containing 4.7 µg of NZMS and 4 units of adenosine deaminase in 50 mM potassium phosphate buffer, pH 7.2, containing 1 mM EDTA (buffer A), is added 200 µl of *S*-Adenosyl-L-homocysteine (500 µM) containing 250 µM DTNB in buffer A. The reaction mixture is incubated at 37°C for 2 minutes. Hydrolytic activity is monitored at 412 nm continuously using a diode array UV spectrophotometer. Enzyme activity is defined as the amount of enzyme that can hydrolyze 1 µmol of *S*-Adenosyl-L-homocysteine/minute (Yuan, C-S et al. (1996) J. Biol. Chem. 271:28009-28015).

NZMS hydrolytic activity is measured in the synthetic direction as the production of S-adenosyl homocysteine using 3-deazaadenosine as a substrate (Sganga, M.W. et al. supra). Briefly, NZMS is incubated in a 100 µl volume containing 0.1 mM 3-deazaadenosine, 5 mM homocysteine, 20

mM Hepes (pH 7.2). The assay mixture is incubated at 37°C for 15 minutes. The reaction is terminated by the addition of 10 µl of 3 M perchloric acid. After incubation on ice for 15 minutes, the mixture is centrifuged for 5 minutes at 18,000 x *g* in a microcentrifuge at 4°C. The supernatant is removed, neutralized by the addition of 1 M potassium carbonate, and centrifuged again. A 50 µl aliquot of supernatant is then chromatographed on an Altex Ultrasphere ODS column (5 µm particles, 4.6 x 250 mm) by isocratic elution with 0.2 M ammonium dihydrogen phosphate (Aldrich) at a flow rate of 1 ml/min. Protein is determined by the bicinchoninic acid assay (Pierce).

Alternatively, NZMS hydrolyase activity can be measured in the synthetic direction by a TLC method (Hershfield, M.S. et al. (1979) J. Biol. Chem. 254:22-25). In a preincubation step, 50 µM [8-^{14}C]adenosine is incubated with 5 molar equivalents of NAD^+ for 15 minutes at 22°C. Assay samples containing NZMS in a 50 µl final volume of 50 mM potassium phosphate buffer, pH 7.4, 1 mM DTT, and 5 mM homocysteine, are mixed with the preincubated [8-^{14}C]adenosine/NAD^+ to initiate the reaction. The reaction is incubated at 37°C, and 1 µl samples are spotted on TLC plates at 5 minute intervals for 30 minutes. The chromatograms are developed in butanol-1/glacial acetic acid/water (12:3:5, v/v) and dried. Standards are used to identify substrate and products under ultraviolet light. The complete spots containing [^{14}C]adenosine and [^{14}C]SAH are then detected by exposing x-ray film to the TLC plate. The radiolabeled substrate and product are then cut from the chromatograms and counted by liquid scintillation spectrometry. Specific activity of the enzyme is determined from the linear least squares slopes of the product *vs* time plots and the milligrams of protein in the sample (Bethin, K.E. et al. (1995) J. Biol. Chem. 270:20698-20702).

XVIII. Identification of NZMS Agonists and Antagonists

Agonists or antagonists of NZMS activation or inhibition may be tested using the assay described in section XVII. Agonists cause an increase in NZMS activity and antagonists cause a decrease in NZMS activity.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
8159895	1	8159895CD1	19	8159895CB1
2497773	2	2497773CD1	20	2497773CB1
354561	3	354561CD1	21	354561CB1
7484682	4	7484682CD1	22	7484682CB1
7485253	5	7485253CD1	23	7485253CB1
2397473	6	2397473CD1	24	2397473CB1
7485243	7	7485243CD1	25	7485243CB1
2199285	8	2199285CD1	26	2199285CB1
2448021	9	2448021CD1	27	2448021CB1
3187209	10	3187209CD1	28	3187209CB1
4507128	11	4507128CD1	29	4507128CB1
5519834	12	5519834CD1	30	5519834CB1
2215017	13	2215017CD1	31	2215017CB1
7484731	14	7484731CD1	32	7484731CB1
3927361	15	3927361CD1	33	3927361CB1
6542758	16	6542758CD1	34	6542758CB1
3188878	17	3188878CD1	35	3188878CB1
7500488	18	7500488CD1	36	7500488CB1

Table 2

Polypep-tide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	8159895CD1	g179793	6.5E-87	[Homo sapiens] carbonic anhydrase I (EC 4.2.1.1) Lowe, N., et al. (1990) Gene 93:277-283 Structure and methylation patterns of the gene encoding human carbonic anhydrase I.
2	2497773CD1	g1263164	1.3E-168	[Rattus norvegicus] cysteine sulfinate decarboxylase Reymond, I., et al. (1996) Biochim. Biophys. Acta 1307:152-156 Molecular cloning and sequence analysis of the cDNA encoding rat liver cysteine sulfinate decarboxylase (CSD).
3	354561CD1	g10580053	6E-29	Dihydrodipicolinate synthase; DapA [Halobacterium sp. NRC-1]
4	7484682CD1	g1185363	2.1E-230	[Rattus norvegicus] S-adenosyl-L-homocysteine hydrolase Merta, A. et al. (1995) Eur. J. Biochem. 229:575-582 The gene and pseudogenes of rat S-adenosyl-L-homocysteine hydrolase.
5	7485253CD1	g8346547	6.4E-48	[Arabidopsis thaliana] asparaginase
6	2397473CD1	g6478485	1.2E-174	[Mus musculus] peroxisomal long chain acyl-CoA thioesterase Ib Hunt, M.C. et al (1999) J. Biol. Chem. 274:34317-34326 Peroxisome proliferator-induced long chain acyl-CoA thioesterases comprise a highly conserved novel multi-gene family involved in lipid metabolism.
7	7485243CD1	g2745760	5.3E-44	[Saguinus oedipus] ribonuclease k6 precursor
8	2199285CD1	g2414618	1.9E-116	[Schizosaccharomyces pombe] ribonuclease II RNB family protein; dis3-like
9	2448021CD1	g1237213	1.8E-68	[Homo sapiens] glyoxalase II Ridderstrom, M., et al. (1996) J. Biol. Chem. 271:319-323 Molecular cloning, heterologous expression, and characterization of human glyoxalase II.
10	3187209CD1	g179077	1.9E-144	[Homo sapiens] arylsulfatase B precursor (EC 3.1.6.1) Peters, C., et al. (1990) J. Biol. Chem. 265:3374-3381 Phylogenetic conservation of arylsulfatases. cDNA cloning and expression of human arylsulfatase B.

Table 2

Polypep-tide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
11	4507128CD1	g2766161	0	[Mus musculus] alpha-D-mannosidase Hiramoto, S., et al. (1997) Biochem. Biophys. Res. Commun. 241:439-445 Stage-specific expression of a mouse homologue of the porcine 135kDa alpha-D-mannosidase (MAN2B2) in type A spermatogonia.
12	5519834CD1	g603768	4.9E-66	[Bacillus subtilis] HutI protein, imidazolone-5-propionate hydrolase
13	2215017CD1	g258856	3.1E-144	[Felis catus] arylsulfatase B, ARSB Jackson, C.E. et al. (1992) Genomics 14:403-411 Feline arylsulfatase B (ARSB): isolation and expression of the cDNA, comparison with human ARSB, and gene localization to feline chromosome A1.
14	7484731CD1	g178279	5.6E-83	[Homo sapiens] S-adenosylhomocysteine hydrolase Coulter-Karis, D.E. et al. (1989) Ann. Hum. Genet. 53:169-175 Sequence of full length cDNA for human S-adenosylhomocysteine hydrolase.
15	3927361CD1	g4105619	8.6E-144	[Mus musculus] SPAF (spermatogenesis associated factor, AAA ATPase family) Liu, Y. et al. (2000) Oncogene 2000 19:1579-1588 SPAF, a new AAA-protein specific to early spermatogenesis and malignant conversion.
17	3188878CD1	g2463026	8.1E-41	[Drosophila melanogaster] PRUNE protein Timmons, L. and Shearn, A. (1996) Genetics 144:1589-1600 Germline transformation using a prune cDNA rescues prune/killer of prune lethality and the prune eye color phenotype in Drosophila.

Table 2

Polypep-tide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
18	7500488CD1	g12655792	2.1E-148	[Homo sapiens] prune (neural development) protein
		372288|SPAC2F3.11	5.8E-20	[Schizosaccharomyces pombe] Putative exopolyphosphatase
		10519|PPX1	9.1E-17	[Saccharomyces cerevisiae] [Other phosphatase; Hydrolase] [Cytoplasmic] Exopolyphosphatase, soluble enzyme that degrades polyphosphate chains of all lengths, with a preference for those of 250 residues

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	8159895CD1	262	S30 S44 S49 S56 S88 S100 S126 S131 S167 T178 Y41	N76 N218	Eukaryotic-type carbonic anhydrase: W6-F261	HMMER_PFAM
					Eukaryotic-type carbonic anhydrases signature BL00162: W17-P47, Y52-T74, Y89-N125, K128-G152, D191-Q223, R228-F261	BLIMPS_BLOCKS
					Eukaryotic-type carbonic anhydrases signature euk_co2_anhydrase.prf: G82-A143	PROFILESCAN
					CARBONIC ANHYDRASE DEHYDRATASE LYASE CARBONATE ZINC PRECURSOR SIGNAL PROTEIN GLYCOPROTEIN PD000865: E10-F261	BLAST_PRODOM
					CARBONIC ANHYDRASE DM00356|JN0836|25-261: D25-F261 DM00356|JN0835|25-261: D25-F261 DM00356|P48282|24-260: D25-F261 DM00356|P00918|23-258: G26-F261	BLAST_DOMO
					Eukaryotic-type carbonic anhydrases signature: S106-V122	MOTIFS
2	2497773CD1	502	S4 S90 S188 S202 S292 S339 S366 S429 S459 S484 T195 T377 Y356 Y418		Pyridoxal-dependent decarboxylase conserved domain: P58-I426	HMMER_PFAM
					DDC/GAD/HDC/TyrDC pyridoxal-phosphate attachment site BL00392: W278-G287	BLIMPS_BLOCKS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					DECARBOXYLASE LYASE PYRIDOXAL PHOSPHATE MULTIGENE FAMILY DOPA GLUTAMATE ACID AROMATIC L-AMINO ACID PD001960: P58-F423	BLAST_PRODOM
					GLUTAMATE DECARBOXYLASE PD114206: E27-N121	BLAST_PRODOM
					DDC / GAD / HDC / TYRDC PYRIDOXAL-PHOSPHATE ATTACHMENT SITE DM00568|S55689|5-478: D22-P466 DM00568|P14748|101-591: A23-K500 DM00568|Q05329|95-582: F29-K500 DM00568|JH0827|84-575: E27-M502	BLAST_DOMO
					DDC / GAD / HDC / TyrDC pyridoxal-phosphate attachment site: S307-K328	MOTIFS
3	354561CD1	281	S9 S77 S87 S110 S198 T85 T174 T212		Dihydrodipicolinate synthetase family domain: V52-L281	HMMER-PFAM
					Dihydrodipicolinate synthetase signature 2: Y168-S198	MOTIFS
					Dihydrodipicolinate synthetase signature: V166-G225	ProfileScan
					Transmembrane domain: L220-A248 N-terminus is non-cytosolic	TMAP
					Dihydrodipicolinate synthase signatures: G37-F46, V52-R104, L105-T155, P164-L186, G222-G247	BLIMPS-BLOCKS
					Dihydrodipicolinate synthase signatures: P68-R89, R104-T122, P138-Y154, I163-V180	BLIMPS-PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Lyase, dihydrodipicolinate synthase, lysine biosynthesis PD001859: G71-I273	BLAST-PRODOM
					Dihydrodipicolinate synthetase: DM00937|Q04796|1-289: V33-I273	BLAST-DOMO
					Dihydrodipicolinate synthetase: DM00937|P40109|11-300: V42-T212 DM00937|P43797|5-297: A36-I273 DM00937|Q07607|1-290: G37-L281	BLAST-DOMO
4	7484682CD1	433	S2 S83 S188 S199 T106 T208 T243 T261 T262 T375 T411 Y110	N181	AdoHcyase: K8-H376	HMMER_PFAM
					TMAP: A48-W76, N-terminus is cytosolic	TMAP
					S-adenosyl-L-homocysteine hydrolase proteins BL00738: Y7-K46, G47-G71, A72-E109, N126-H140, G152-K174, N179-V210, A238-E259, V260-L312, I337-W374, V384-Y433	BLIMPS_BLOCKS
					HYDROLASE ADENOSYL-HOMOCYSTEINASE ADOHCYASE NAD PD001319: K8-P144, K142-I201 PD000699: D202-D308 PD149655: E306-H376 PD149849: Q145-L200	BLAST_PRODOM
					S-ADENOSYL-L-HOMOCYSTEINE HYDROLASE DM01437|JC2480|2-433: S2-Y433	BLAST_DOMO
					S-ADENOSYL-L-HOMOCYSTEINE HYDROLASE DM01437|P27604|3-436: K4-Y433	BLAST_DOMO
					S-ADENOSYL-L-HOMOCYSTEINE HYDROLASE DM01437|S50546|3-449: Y7-Y433	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					S-ADENOSYL-L-HOMOCYSTEINE HYDROLASE DM01437|P28183|1-462: Y143-Y433, Y7-T136	BLAST_DOMO
					S-adenosyl-L-homocysteine hydrolase signature 1 and 2: C79-A93, G214-A230	MOTIFS
5	7485253CD1	308	S43 S80 T71 T141 T243 T303		Asparaginase: M1-T302	HMMER_PFAM
					HYDROLASE N4- PRECURSOR PD02894: I81-A126, G178-D210	BLIMPS_PRODOM
					PRECURSOR HYDROLASE SIGNAL N'-(beta-N-ACETYLGLUCOSAMINYL)-L-ASPARAGINASE GLYCOSYLASPARAGINASE ASPARTYLGLUCOSAMINIDASE AMIDASE AGA L-ASPARAGINASE N4-N-ACETYL-□ PD005819: V5-H114, T141-D300, A30-K152	BLAST_PRODOM
					N'(beta-N-ACETYLGLUCOSAMINYL)-L-ASPARAGINASE PRECURSOR EC 3.5.1.26 GLYCOSYLASPARAGINASE ASPARTYLGLUCOSAMINIDASE N'-N-ACETYL-beta- GLUCOSAMINYL-L-ASPARAGINE AMIDASE AGA SIGNAL HYDROLASE PERIP PD114843: V28-G130	BLAST_DOMO
					GLYCOSYLASPARAGINASE CHAIN DM07808|P50287|40-314: L38-D301 DM07808|P20933|19-345: A30-G130, Q158-A276	BLAST_DOMO
6	2397473CD1	421	S37 S138 S339 T33 T337 T413	N202 N247 N255	TRANSMEMBRANE DOMAIN: P158-L186, P219-N247 N-terminus is cytosolic	TMAP

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					ACYLCOA ACID HYDROLASE PROTEIN THIOESTERASE KAN1 BILE COA: AMINO NACYLTRANSFERASE PD006914: L5-G411	BLAST_PRODOM
					ACID; KAN-1; COA; AMINO; DM05400|S59131|1-420: N16-L410 DM05400|A53965|1-418: N16-L410	BLAST_DOMO
7	7485243CD1	155	S27 S113 T62 T85		signal_cleavage: M1-A22	SPSCAN
					Signal cleavage: A28	HMMER
					Pancreatic ribonucleases: G32-V155	HMMER_PFAM
					TRANSMEMBRANE DOMAIN: A4-P25 N-terminus is non-cytosolic	TMAP
					Pancreatic ribonuclease BL00127: S36-Q45, C51-K95, S105-P148	BLIMPS_BLOCKS
					Pancreatic ribonuclease family signature rnase_pancreatic.prf: P46-K91	PROFILESCAN
					Pancreatic ribonuclease family signature PR00794: C51-T70, F71-C90, N96-G114, P117-Q139	BLIMPS_PRINTS
					HYDROLASE NUCLEASE ENDONUCLEASE RIBONUCLEASE RNASE GLYCOPROTEIN PRECURSOR SIGNAL PANCREATIC A PD000535: Q37-D152	BLAST_PRODOM
					RIBONUCLEASE PANCREATIC RNASE A HYDROLASE NUCLEASE ENDONUCLEASE GLYCOPROTEIN PD152095: E74-V155	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PANCREATIC RIBONUCLEASE FAMILY DM00621|JC2034|5-126: M33-V154 DM00621|P08904|5-126: M33-K153 DM00621|I61900|32-158: S36-D152 DM00621|P47778|32-159: T34-V155	BLAST_DOMO
					Pancreatic ribonuclease family signature C65-F71	MOTIFS
8	2199285CD1	885	S2 S31 S41 S48 S55 S154 S157 S173 S194 S198 S225 S232 S244 S269 S369 S395 S421 S459 S499 S503 S521 S578 S746 S838 S875 T15 T67 T210 T386 T505 T647 Y137	N39 N641	RNB-like proteins domain: A291-C676	HMMER_PFAM
					Ribonuclease II family proteins signature BL01175: R371-P398, S421-P457, L581-N590, Y696-R713	BLIMPS_BLOCKS
					PROTEIN HYDROLASE NUCLEASE VACB HOMOLOG RIBONUCLEASE II DIS3 NUCLEAR EXORIBONUCLEASE PD003098: R292-Y674	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					do VACB; II; EXORIBONUCLEASE; DM01952|Q09568|119-795: G201-Q863 DM01952|P21499|127-640: K321-Y674 DM01952|P37202|327-965: D220-A801 DM01952|P44907|135-730: E366-D633	BLAST_DOMO
					Ribonuclease II family signature: H688-H712	MOTIFS
9	2448021CD1	282	S143 S155 S177 T86 T213 T235		Metallo-beta-lactamase superfamily domain: P7-H172	HMMER_PFAM
					PROBABLE HYDROXYACYLGLUTATHIONE HYDROLASE EC 3.1.2.6 GLYOXALASE II GLX PD082397: D29-G92 (P-value = 7/8e-09)	BLAST_PRODOM
					do RNH; ATP; HI1663; SYNTHASE; DM02001|P05446|81-189: A78-N189 DM02001|Q08889|77-182: S84-P187 DM02001|G64113|76-181: T107-V185	BLAST_DOMO
10	3187209CD1	576	S275 S363 S390 S486 S501 S542 S553 T108 T166 T220 T324 T383 T466 T536 Y146	N134 N283 N295 N408 N474 N504	signal_cleavage: M1-G24	SPSCAN
					Signal Peptide: M4-L21, M4-G24, M1-G24, M1-E28	HMMER
					Sulfatase domain: P53-P479	HMMER_PFAM
					Transmembrane domain: M4-S22 N-terminus is non-cytosolic	TMAP
					Sulfatases proteins signature BL00523: P53-G69, C99-K110, G145-H155, P235-H246, L277-G306, D356-E366, L472-E481	BLIMPS_BLOCKS
					Sulfatases signature: Q126-G175	PROFILESCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					HYDROLASE ARYLSULFATASE PRECURSOR SIGNAL GLYCOPROTEIN LYSOSOME PROTEIN SULPHOHYDROLASE MUCOPOLYSACCHARIDOSIS SULFATASE PD001700: P53-Y259, T216-P490	BLAST_PRODOM
					ARYLSULFATASE B PRECURSOR ASB NACETYLGALACTOSAMINE 4SULFATASE G4S HYDROLASE SIGNAL GLYCOPROTEIN PD037102: H397-W531	BLAST_PRODOM
					ARYLSULFATASE HYDROLASE PRECURSOR ARYLSULFATE SULPHOHYDROLASE ARS SIGNAL GLYCOPROTEIN EXTRACELLULAR MATRIX PD035731: L35-M194	BLAST_PRODOM
					SIMILAR TO ARYLSULFATASE B PD023029: I300-W411	BLAST_PRODOM
					SULFATASES DM01026|P33727|44-518: P53-P513 DM01026|P34059|28-486: P53-A421 DM01026|P50473|63-522: S51-Y415 DM01026|P15289|18-477: P53-G416	BLAST_DOMO
					Sulfatases signature 1 P97-G109	MOTIFS
					Sulfatases signature 2 G145-H155	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11	4507128CD1	1009	S35 S48 S59 S174 S199 S201 S225 S367 S376 S428 S522 S536 S574 S601 S607 S775 S968 T44 T120 T324 T338 T363 T424 T573 T610 T810 Y247 Y350 Y620	N226 N249 N294 N336 N516 N608 N670 N675 N748 N808 N812 N890	Signal Peptide: M1-S23	HMMER
					Glycosyl hydrolases family: M1-T614	HMMER_PFAM
					TMAP: V479-V500 D813-T831 N-terminus is cytosolic	TMAP
					Glycosyl hydrolases family signatures PF01074: I27-M49, V96-G141, P144-V193, F277-N310, D346-R373, Q418-L439, S632-K641, S750-E790	BLIMPS_PFAM
					ALPHAMANNOSIDASE HYDROLASE GLYCOSIDASE GLYCOPROTEIN LYSOSOMAL MANNOSYLOLIGOSACCHARIDE 6ALPHAMANNOSIDASE MAN TRANSMEMBRANE SIGNALANCHOR PD003951: I27-A473, V479-V613 PD003984: S632-L961	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					EPIDIDYMISSPECIFIC ALPHAMANNOSIDASE ALPHADMANNOSIDE MANNOHYDROLASE PROTEIN HYDROLASE GLYCOSIDASE MANNOSIDASE ALPHA B2 PD043751: W798-Q1008	BLAST_PRODOM
					LUMENAL DOMAIN DM02462|P34098|1-566: I27-F224, Q243-A546 DM02462|P49641|123-725: A24-M450, D469-L535 DM02462|P27046|122-725: P26-M450, K430-I542 DM02462|JC2200|4-536: L4-R373, P248-G442	BLAST_DOMO
					Cell attachment sequence: R977-D979	MOTIFS
					ATP/GTP-binding site motif A (P-loop): A762-S769	MOTIFS
12	5519834CD1	426	S75 S242 S282 S358 S396 S397 T67 T108 T130 T151 T199 T243 T297	N395	Urease domain: T367-I393	HMMER_PFAM
					HYDROLASE IMIDAZOLONE5PROPIONATE PROTEIN IMIDAZOLONEPROPIONASE HISTIDINE METABOLISM COAGGREGATION MEDIATING ADHESIN SCAA PD014595: M145-S358	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					HYDROLASE IMIDAZOLONE5PROPIONATE IMIDAZOLONEPROPIONASE HISTIDINE METABOLISM COSMID T12A2 PD023469: I79-R136 PD038142: E361-I423	BLAST_PRODOM
					ATP/GTP-binding site motif A (P-loop): A370-S377	MOTIFS
13	2215017CD1	583	S3 S81 S258 S377 S397 S471 S539 S580 T30 T116 T174 T282 T331 T354 Y154	N290 N302 N480 N510	Sulfatase: P61-P485	HMMER_PFAM
					Sulfatases proteins. BL00523: D363-G373, L478-E487, P61-G77, C107-R118, G153-H163, P242-H253, M284-G313	BLIMPS_BLOCKS
					Sulfatases signatures sulfatase 2: P135-G183	PROFILESCAN
					HYDROLASE ARYLSULFATASE PRECURSOR SIGNAL GLYCOPROTEIN LYSOSOME PROTEIN SULPHOHYDROLASE MUCOPOLYSACCHARIDOSIS SULFATASE PD001700: P61-Y266, S236-W425, L478-D497	BLAST_PRODOM
					ARYLSULFATASE B PRECURSOR ASB N-ACETYLGALACTOSAMINE 4-SULFATASE G4S HYDROLASE SIGNAL GLYCOPROTEIN PD037102: W425-W537	BLAST_PRODOM
					ARYLSULFATASE HYDROLASE PRECURSOR ARYLSULFATE SULPHOHYDROLASE ARS SIGNAL GLYCOPROTEIN EXTRACELLULAR MATRIX PD035731: T58-V202	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					SIMILAR TO ARYLSULFATASE B PD023029: S310-H412, A431-D444	BLAST_PRODOM
					SULFATASES DM01026|P33727|44-518: K59-E521 DM01026|P34059|28-486: K59-A428 DM01026|P15289|18-477: K59-L502 DM01026|P50473|63-522: T58-H404	BLAST_DOMO
					Sulfatases signature 1 P105-G117	MOTIFS
					Sulfatases signature 2 G153-H163	MOTIFS
14	7484731CD1	395	S27 S68 S70 S237 T224 T281 T299 T300 T384	N220	S-adenosyl-L-homocysteine hydrolase: M206-V357, K76-A140	HMMER_PFAM
					Transmembrane domain: L122-T150 M175-L192 D240-Y260 N is non-cytosolic	TMAP
					S-adenosyl-L-homocysteine hydrolase proteins BL00738: I217-V248, A276-E297, V298-L350, S75-E114, G115-G139	BLIMPS_BLOCKS
					HYDROLASE AdoHcyase NAD ONE-CARBON METABOLISM S-ADENOSYL-L-HOMO-CYSTEINE PUTATIVE PD001319: K76-K205, K205-I239	BLAST_PRODOM
					NAD DEHYDROGENASE OXIDOREDUCTASE HYDROLASE AdoHcyase ONE-CARBON METABOLISM PROTEIN S-ADENOSYL-L-HOMOCYSTEINE PD000699: G233-V343	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					S-ADENOSYL-L-HOMOCYSTEINE HYDROLASE DM01437 JC2480|2-433: N200-K362, S70-I211, K352-G392 S50546|3-449: K205-K362, K76-K205, E344-G392 P27604|3-436: K205-D365, K76-K205, E342-G392 P35007|9-484: K205-P354, E342-L391, K76-H203	BLAST_DOMO
					S-adenosyl-L-homocysteine hydrolase signature 2 G252-A268	MOTIFS
15	3927361CD1	503	S33 S37 S39 S89 S90 S113 S199 S258 S260 S272 S277 S286 T32 T297 T311 T343 T405 T425 T470 Y148	N50 N98 N103	Signal_cleavage: M1-S33	SPSCAN
					ATPases associated with various cellular activities: G334-L502	HMMER_PFAM
					AAA protein family protein BL00674 Q332-A353, T405-N451, G483-L502	BLIMPS_BLOCKS
					PROTEIN ATPBINDING PROTEASE SUBUNIT HOMOLOG REPEAT CELL DIVISION ATPDEPENDENT NUCLEAR PD000092: I375-A496, G334-A358	BLAST_PRODOM
					AAA-PROTEIN FAMILY DM00024 S64785|200-363: E367-R436, K333-S370 P46464|456-616: P373-L435, N312-K371 P40340|408-571: E367-R436, M320-A353 Q07590|481-641: T330-K371, P373-L435	BLAST_DOMO
					AAA-protein family signature V419-R437	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					ATP/GTP-binding site motif A (P-loop) G339-T346	MOTIFS
16	6542758CD1	165	S79 S102 S120 T81		Signal_cleavage M44-L71	HMMER
					TMAP: A48-P68 V126-F154 N terminus cytosolic	TMAP
					Eukaryotic thiol (cysteine) proteases active sites thiol_protease_cys.prf: E4-C52	PROFILESCAN
					Eukaryotic thiol (cysteine) proteases cysteine active site Q21-A32	MOTIFS
17	3188878CD1	453	S111 S347 S365 S399 S414 S451 T46 T115 T153 T191 T291 T311		TMAP: R16-K44 N terminus: non-cytosolic	TMAP
					PRUNE EXOPOLYPHOSPHATASE METAPHOSPHATASE PROTEIN HYDROLASE GENE PUTATIVE XPP PD011764: E50-G245, R16-E154, D213-L359	BLAST_PRODOM
					Leucine zipper pattern L157-L178 L164-L185	MOTIFS
18	7500488CD1	400	S111 S294 S312		Signal_cleavage: M1-A48	SPSCAN
					DHHA2 domain: F215-L306	HMMER_PFAM
					PRUNE EXOPOLYPHOSPHATASE METAPHOSPHATASE PROTEIN HYDROLASE GENE PUTATIVE XPP PD011764: E50-G245, R16-E154, K236-L306	BLAST_PRODOM
					Leucine zipper pattern: L157-L178, L164-L185	MOTIFS
					Cell attachment sequence: R66-D68	MOTIFS

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
19/8159895CB1/1023	1-371, 1-447, 1-532, 1-549, 1-556, 1-570, 27-409, 31-639, 203-468, 209-690, 209-694, 231-743, 236-552, 261-686, 312-985, 332-735, 359-743, 529-1023, 547-1023, 683-834, 988-1017
20/2497773CB1/1848	1-243, 1-302, 1-510, 1-664, 151-814, 249-758, 249-894, 279-898, 441-1131, 496-1170, 502-1056, 504-1031, 521-1114, 535-1191, 539-976, 539-990, 540-1081, 547-1042, 550-1087, 639-1325, 640-1273, 838-1529, 924-1467, 1085-1493, 1158-1579, 1202-1649, 1264-1848, 1325-1848, 1394-1568
21/354561CB1/1336	1-309, 1-434, 6-647, 35-630, 46-268, 75-617, 75-672, 90-337, 90-506, 274-485, 369-1014, 646-1272, 703-916, 976-1336, 1027-1198, 1064-1198, 1071-1198, 1198-1336
22/7484682CB1/1302	1-1302
23/7485253CB1/1428	1-260, 87-242, 91-656, 97-550, 104-722, 105-620, 107-348, 107-607, 107-636, 107-660, 107-692, 107-758, 108-752, 115-630, 137-792, 149-937, 156-767, 268-860, 272-801, 317-901, 340-919, 364-975, 371-955, 378-897, 378-919, 383-985, 388-952, 411-933, 418-818, 436-698, 438-958, 438-1017, 439-957, 449-899, 456-988, 458-1101, 468-968, 470-1079, 500-951, 502-804, 519-1079, 528-1146, 533-1094, 536-1137, 549-1111, 552-820, 554-1132, 556-1006, 565-1105, 572-1173, 576-742, 609-1135, 619-813, 623-1330, 626-1088, 627-1313, 630-919, 631-873, 639-1252, 654-936, 654-1008, 663-854, 664-850, 679-855, 699-833, 699-1134, 712-1117, 716-891, 723-853, 753-1030, 758-1282, 799-1411, 816-1342, 832-1330, 871-1380, 874-1428, 909-1160, 911-1404, 926-1425, 927-1428, 961-1420, 967-1323, 1056-1428, 1272-1301
24/2397473CB1/1393	1-457, 191-420, 192-658, 192-660, 224-889, 432-753, 432-756, 450-735, 476-1111, 504-846, 565-796, 565-1039, 579-1232, 591-867, 612-1284, 659-1376, 697-997, 711-1363, 788-1393, 811-1365, 864-1312, 873-1393, 886-1153, 910-1148, 921-1375, 924-1380, 942-1393, 960-1376, 960-1377, 964-1331, 982-1376, 1031-1379, 1115-1378, 1124-1303, 1124-1363, 1186-1339
25/7485243CB1/567	1-259, 100-567, 300-373, 300-472

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
26/2199285CB1/3519	1-759, 6-43, 164-649, 168-465, 176-349, 176-416, 184-429, 186-499, 204-463, 204-755, 497-759, 497-1079, 500-1133, 598-1230, 611-779, 734-1051, 739-1321, 800-1372, 833-1366, 892-1473, 894-1189, 908-1452, 915-1423, 915-1511, 946-1193, 1137-1411, 1161-1742, 1187-1618, 1202-1821, 1296-1624, 1300-1554, 1338-1609, 1338-1750, 1366-1845, 1379-1642, 1434-1731, 1460-1709, 1466-2047, 1496-2009, 1540-2162, 1559-2195, 1589-2000, 1611-2259, 1635-1979, 1637-2167, 1683-1929, 1695-2206, 1704-2300, 1739-2250, 1743-1888, 1745-2001, 1805-2072, 1849-2114, 1902-2587, 1992-2379, 1994-2548, 2052-2559, 2068-2339, 2079-2301, 2221-2463, 2233-2492, 2256-2937, 2270-2505, 2270-2793, 2287-2819, 2418-3014, 2484-2746, 2528-2757, 2580-2996, 2606-2795, 2629-2874, 2629-3283, 2633-3117, 2644-3285, 2680-2881, 2689-2965, 2694-2941, 2823-3073, 2823-3282, 2823-3293, 2846-3287, 2860-3101, 2920-3267, 3036-3300, 3036-3301, 3044-3515, 3046-3519, 3072-3302, 3076-3301, 3120-3302, 3167-3517, 3205-3517, 3230-3519
27/2448021CB1/1291	1-249, 1-265, 1-365, 1-592, 1-640, 1-641, 1-643, 1-664, 1-670, 1-691, 1-717, 19-706, 40-549, 53-222, 53-269, 53-274, 53-323, 53-446, 53-459, 53-501, 53-527, 55-381, 55-497, 55-522, 86-227, 86-341, 86-439, 95-308, 110-514, 122-411, 126-453, 139-448, 147-749, 157-435, 335-775, 341-1111, 441-674, 445-743, 456-727, 490-779, 517-677, 548-724, 548-1139, 557-1263, 559-1217, 565-830, 565-858, 567-1269, 580-831, 594-1162, 600-1265, 619-861, 619-885, 619-1187, 629-978, 658-1159, 664-1259, 673-828, 698-964, 699-972, 710-945, 710-950, 711-831, 752-782, 755-831, 761-1247, 763-992, 763-1110, 774-831, 779-1015, 784-1059, 798-831, 803-831, 816-1291, 832-1004, 832-1212, 832-1241, 832-1250, 832-1255, 832-1259, 832-1262, 832-1265, 832-1276, 832-1277, 836-1115, 846-1261, 849-1265, 850-1259, 850-1266, 856-967, 856-1103, 858-1259, 865-1259, 865-1279, 874-1262, 899-1266, 904-1267, 928-1250, 934-1195, 938-1276, 948-1082, 956-1267, 1012-1245, 1012-1269, 1019-1266, 1028-1184, 1033-1259, 1034-1266, 1052-1265, 1087-1257, 1087-1274

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
28/3187209CB1/3072	1-290, 1-609, 1-2368, 11-610, 361-935, 610-2368, 830-1123, 830-1266, 884-1098, 884-1402, 1048-1337, 1072-1629, 1243-1885, 1269-1889, 1275-1820, 1285-1888, 1337-1917, 1337-1919, 1343-1896, 1400-1705, 1402-1992, 1408-2002, 1440-1989, 1489-2020, 1499-2047, 1505-1939, 1512-1949, 1514-2119, 1514-2149, 1517-1760, 1558-1930, 1619-1713, 1619-1841, 1634-1948, 1640-2302, 1642-1874, 1651-1930, 1654-1927, 1677-2413, 1681-2256, 1715-2430, 1729-2406, 1729-2456, 1734-2357, 1737-2016, 1737-2307, 1769-2350, 1812-2078, 1893-2543, 1901-2527, 1953-2633, 1965-2633, 2062-2722, 2063-2617, 2072-2537, 2082-2731, 2089-2763, 2102-2331, 2127-2780, 2143-2785, 2158-2731, 2164-2842, 2183-2722, 2293-2883, 2315-2978, 2473-2662, 2477-3047, 2499-3063, 2499-3072, 2509-2979, 2556-2816, 2633-2902, 2745-3010, 2788-3050, 2813-3042, 2813-3072, 2816-3050, 2887-3063, 2913-3063, 2935-3072, 2955-3057
29/4507128CB1/4117	1-244, 1-445, 4-513, 7-445, 10-644, 17-308, 20-108, 28-359, 28-474, 29-316, 29-496, 30-320, 30-436, 39-493, 39-628, 42-284, 52-605, 142-796, 183-706, 186-591, 275-473, 275-576, 370-780, 417-3055, 453-702, 461-759, 712-1283, 758-1226, 845-1149, 874-1312, 923-1394, 997-1258, 997-1593, 1309-1572, 1402-1968, 1496-1795, 1500-1908, 1624-1924, 1727-2327, 1775-1964, 1802-2201, 1847-2092, 1940-2173, 1943-2146, 1959-2244, 2060-2336, 2069-2519, 2069-2538, 2069-2562, 2103-2371, 2162-2404, 2205-2494, 2205-2691, 2378-2657, 2378-2955, 2411-2878, 2420-2674, 2445-2993, 2481-2925, 2551-2817, 2561-3003, 2689-3083, 2689-3148, 2689-3162, 2722-2981, 2725-2969, 2748-3247, 2819-3359, 2834-3236, 2869-3155, 2869-3525, 2879-3144, 2934-3171, 2943-3200, 2963-3301, 2972-3524, 3015-3306, 3061-3352, 3126-3725, 3127-3330, 3138-3478, 3138-3631, 3166-3601, 3176-3638, 3176-3667, 3193-3648, 3199-3457, 3216-3501, 3221-3500, 3229-3538, 3229-4101, 3239-3559, 3247-3497, 3255-3549, 3255-3607, 3258-3520, 3262-3512, 3294-3524, 3306-3530, 3314-3889, 3320-3656, 3324-3523, 3325-3581,
	3345-3783, 3348-3899, 3351-3484, 3352-3596, 3363-3610, 3363-3612, 3401-3630, 3478-3704, 3478-3770, 3483-3749, 3497-3750, 3500-3762, 3505-3753, 3512-3704, 3522-4117, 3535-3829, 3539-3723, 3596-4117, 3657-3924, 3682-3822
30/5519834CB1/2340	1-409, 1-529, 469-851, 517-851, 651-906, 652-1102, 899-1153, 952-1145, 969-1472, 990-1228, 1062-1651, 1158-1652, 1161-1460, 1316-1607, 1329-1521, 1545-1814, 1604-2255, 1706-2007, 1706-2098, 1830-2314, 1838-2304, 1840-2304, 1846-2301, 1853-2299, 1856-2309, 1857-2309, 1861-2307, 1863-2301, 1874-2281, 1885-2300, 1891-2301, 1897-2320, 1901-2148, 1910-2304, 1917-2301, 1919-2301, 1977-2303, 2009-2307, 2040-2287, 2081-2340, 2131-2301

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
31/2215017CB1/2634	1-1752, 579-1088, 794-1423, 962-1216, 1013-1230, 1013-1278, 1013-1291, 1108-1351, 1108-1382, 1108-1536, 1108-1576, 1108-1580, 1108-1601, 1108-1634, 1108-1653, 1108-1655, 1108-1669, 1108-1677, 1108-1678, 1108-1680, 1108-1683, 1108-1742, 1108-1751, 1109-1451, 1109-1723, 1111-1759, 1186-1807, 1227-1457, 1227-1727, 1258-1807, 1262-1839, 1282-1815, 1285-1535, 1285-1607, 1285-1659, 1285-1689, 1285-1799, 1285-1960, 1287-1784, 1301-1665, 1305-1987, 1308-1537, 1320-1565, 1320-1988, 1328-1909, 1332-2005, 1338-1527, 1338-1994, 1346-1553, 1350-1928, 1366-1615, 1366-1894, 1377-2031, 1386-1983, 1389-1534, 1406-1882, 1409-1995, 1443-1957, 1446-1892, 1455-1726, 1460-2037, 1470-1785, 1470-2108, 1470-2171, 1478-2002, 1478-2053, 1480-1957, 1483-2139, 1485-2170, 1486-2060, 1503-1891, 1522-1701, 1555-2169, 1567-2190, 1596-2140, 1596-2175, 1596-2270, 1613-2174, 1614-2100, 1619-2270, 1634-2189, 1640-1953, 1651-2294, 1658-2128, 1659-2328, 1669-2171, 1679-2144, 1684-2091, 1684-2359, 1688-1904, 1688-2293, 1702-2289, 1702-2291, 1705-1990,
	1705-2144, 1705-2295, 1707-1874, 1709-2055, 1713-2322, 1713-2369, 1715-2353, 1716-2296, 1727-2261, 1727-2295, 1737-1999, 1737-2281, 1749-2271, 1762-2210, 1763-2190, 1764-1997, 1767-2306, 1770-2315, 1771-2341, 1775-2329, 1778-2395, 1782-2313, 1782-2438, 1789-2417, 1790-2336, 1796-2443, 1806-2313, 1821-2070, 1836-2356, 1860-2394, 1864-2241, 1866-2385, 1871-2418, 1871-2419, 1875-2360, 1886-2506, 1891-2458, 1904-2631, 1906-2187, 1906-2463, 1926-2508, 1941-2627, 1942-2612, 1957-2616, 1971-2606, 1986-2603, 1991-2502, 1995-2508, 2002-2634, 2027-2221, 2030-2612, 2031-2174, 2055-2615, 2057-2634, 2059-2244, 2060-2284, 2061-2442, 2072-2634, 2075-2530, 2103-2579, 2140-2634, 2143-2634, 2145-2579, 2151-2453, 2166-2625, 2166-2634, 2179-2622, 2179-2634, 2183-2378, 2210-2623, 2228-2622, 2258-2568, 2258-2588, 2258-2622, 2258-2631, 2259-2616, 2261-2634, 2262-2631, 2290-2634, 2317-2624, 2348-2485, 2362-2575, 2406-2634, 2421-2634
32/7484731CB1/1188	1-1188, 157-426, 157-565, 157-577, 157-615, 159-604, 615-1188, 647-1051, 775-1051
33/3927361CB1/1670	1-1236, 999-1113, 1030-1113, 1113-1140, 1113-1471, 1283-1670
34/6542758CB1/1070	1-551, 49-535, 355-837, 370-541, 586-1070, 591-1061, 609-924, 619-1062, 621-1065, 672-1062, 683-1064, 700-1064, 715-798, 865-1062, 903-1062, 954-1062

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
35/3188878CB1/2000	1-214, 1-417, 6-389, 11-285, 11-303, 11-307, 21-629, 22-137, 30-306, 38-281, 38-576, 45-306, 58-645, 206-306, 462-807, 487-1059, 487-1131, 593-1084, 599-1105, 655-1225, 693-1279, 694-948, 739-1369, 739-1415, 741-948, 787-1608, 794-948, 823-1619, 852-948, 852-1370, 856-1517, 874-1556, 880-1571, 899-948, 899-1446, 904-1594, 920-1423, 936-1515, 938-1501, 945-1278, 947-1574, 950-1674, 953-1349, 957-1408, 984-1662, 1011-1676, 1018-1552, 1025-1602, 1031-1711, 1041-1250, 1063-1588, 1070-1843, 1077-1618, 1085-1611, 1107-1184, 1107-1187, 1107-1220, 1107-1393, 1107-1397, 1107-1471, 1107-1555, 1107-1584, 1107-1600, 1107-1607, 1107-1734, 1128-1336, 1136-1706, 1143-1555, 1154-1677, 1155-1867, 1163-1749, 1173-1544, 1180-1902, 1186-1446, 1193-1892, 1197-1672, 1202-1578, 1213-1794, 1215-1932, 1218-1673, 1219-1672, 1221-1625, 1228-1487, 1232-2000, 1236-1947, 1238-1902, 1240-1673, 1240-1719, 1251-1887, 1252-1794, 1253-1676, 1254-1671, 1258-2000, 1265-1673, 1271-1673, 1273-2000, 1274-1918, 1275-1671, 1278-1693, 1278-1992, 1280-1865, 1283-1738, 1295-1955, 1314-1961,
	1322-1929, 1343-1954, 1353-1867, 1362-1820, 1369-1841, 1394-1915, 1394-1979, 1405-1979, 1431-2000, 1440-1945, 1450-1846, 1451-2000, 1454-2000, 1456-2000, 1469-2000, 1473-1977, 1480-1977, 1500-2000, 1506-2000, 1507-2000, 1529-2000, 1541-2000, 1550-2000, 1573-2000, 1580-2000, 1596-1884, 1620-2000, 1627-2000, 1648-1945, 1670-1930, 1671-2000, 1675-1946, 1723-2000, 1744-2000, 1745-2000, 1747-1990, 1751-2000, 1758-2000, 1759-2000, 1761-1889, 1761-1976, 1761-2000, 1766-1945, 1767-2000, 1768-2000, 1771-2000, 1786-2000, 1789-2000, 1798-2000, 1802-2000, 1809-2000, 1829-2000, 1834-2000, 1836-2000, 1837-2000, 1850-2000, 1862-2000, 1889-2000, 1914-2000, 1919-2000, 1928-1974, 1957-2000, 1964-2000, 1970-2000
36/7500488CB1/2559	1-730, 513-860, 513-892, 513-901, 513-915, 513-929, 513-2353, 523-815, 523-819, 533-1141, 534-649, 551-1088, 570-1157, 575-1177, 575-1295, 575-1302, 575-1416, 577-1230, 590-818, 616-929, 752-1270, 770-819, 785-1046, 785-1302, 860-1304, 907-1103, 907-1149, 907-1321, 907-1442, 974-1319, 981-1425, 1004-1228, 1006-1458, 1060-1234, 1060-1236, 1206-1749, 1253-1750, 1364-1609, 1364-1824, 1411-1937, 1440-2297, 1452-1953, 1455-2086, 1455-2297, 1464-2297, 1467-2193, 1469-2164, 1481-1687, 1490-2059, 1492-2286, 1496-1908, 1501-2157, 1507-2030, 1508-2220, 1516-2102, 1522-2258, 1526-1896, 1533-2225, 1535-2295, 1542-1795, 1542-2293, 1542-2295, 1547-2297, 1550-2025, 1552-2297, 1555-1930, 1555-2278, 1562-2033, 1566-2147, 1568-2284, 1571-2026, 1572-2025, 1574-1978, 1582-1839, 1591-2254, 1591-2307, 1592-2297, 1593-2026, 1593-2070, 1604-2240, 1605-2027, 1605-2147, 1606-2029, 1607-2024, 1616-2100, 1618-2026, 1626-2158, 1627-2271, 1628-2024, 1633-2046, 1633-2075, 1633-2217, 1637-2089, 1648-2304, 1667-2314, 1672-2030, 1675-2282, 1694-2027, 1696-2306, 1697-2026,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1706-2220, 1719-2017, 1720-2122, 1722-2194, 1741-1998, 1747-2268, 1758-2332, 1793-2297, 1805-2199, 1816-2182, 1826-2094, 1826-2330, 1833-2330, 1859-2559, 1913-2169, 1913-2291, 2001-2297, 2023-2283, 2028-2299, 2100-2343, 2112-2353, 2114-2242, 2114-2329, 2119-2297, 2120-2353

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
19	8159895CB1	MIXDTME02
20	2497773CB1	ADRETUT05
21	354561CB1	CONNTUT05
23	7485253CB1	UTREDIT07
24	2397473CB1	THP1AZT01
25	7485243CB1	PROSNOT06
26	2199285CB1	PITUDIR01
27	2448021CB1	LUNGNOT23
28	3187209CB1	SMCBUNT01
29	4507128CB1	BRAFTUE03
30	5519834CB1	THYMNOT05
31	2215017CB1	SINTFET03
32	7484731CB1	THYMNOT05
33	3927361CB1	KIDNNOT19
34	6542758CB1	LNODNON02
35	3188878CB1	BRABDIR01
36	7500488CB1	BRABDIR01

Table 6

Library	Vector	Library Description
ADRETUT05	pINCY	Library was constructed using RNA isolated from adrenal tumor tissue removed from a 52-year-old Caucasian female during a unilateral adrenalectomy. Pathology indicated a pheochromocytoma.
BRABDIR01	pINCY	Library was constructed using RNA isolated from diseased cerebellum tissue removed from the brain of a 57-year-old Caucasian male, who died from a cerebrovascular accident. Patient history included Huntington's disease,emphysema, and tobacco abuse.
BRAFTUE03	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from brain tumor tissue removed from the left frontal lobe of a 40-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated grade 4 gemistocytic astrocytoma. The patient presented with coma, epilepsy, and incontinence of urine and stool, type II diabetes, abulia, and paralysis. Patient history included chronic nephritis and cesarean delivery. Patient medications included Decadron and phenytoin sodium.
CONNTUT05	pINCY	Library was constructed using RNA isolated from tumorous skull soft tissue removed from a 34-year-old Caucasian female during skull lesion excision. Pathology indicated grade 3 ependymoma forming an implant in the dermis and subcutis associated with dense fibrosis. Patient history included seizures, bone cancer, and brain cancer. Surgeries included cranioplasty and cerebral meninges lesion excision, and treatment included whole brain radiation. Family history included anxiety and depression.
KIDNNOT19	pINCY	Library was constructed using RNA isolated from kidney tissue removed a 65-year-old Caucasian male during an exploratory laparotomy and nephroureterectomy. Pathology for the associated tumor tissue indicated a grade 1 renal cell carcinoma within the upper pole of the left kidney. Patient history included malignant melanoma of the abdominal skin, benign neoplasm of colon, cerebrovascular disease, and umbilical hernia. Family history included myocardial infarction, atherosclerotic coronary artery disease, cerebrovascular disease, prostate cancer, myocardial infarction, and atherosclerotic coronary artery disease.
LNODNON02	pINCY	This normalized lymph node tissue library was constructed from .56 million independent clones from a lymph node tissue library. Starting RNA was made from lymph node tissue removed from a 16-month-old Caucasian male who died from head trauma. Serologies were negative. Patient history included bronchitis. Patient medications included Dopamine, Dobutamine, Vancomycin, Vasopressin, Proventil, and Atarax. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9932 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.

Table 6

Library	Vector	Library Description
LUNGNOT23	pINCY	Library was constructed using RNA isolated from left lobe lung tissue removed from a 58-year-old Caucasian male. Pathology for the associated tumor tissue indicated metastatic grade 3 (of 4) osteosarcoma. Patient history included soft tissue cancer, secondary cancer of the lung, prostate cancer, and an acute duodenal ulcer with hemorrhage. Family history included prostate cancer, breast cancer, and acute leukemia.
MIXDTME02	PBK-CMV	This 5' biased random primed library was constructed using pooled cDNA from five donors. cDNA was generated using mRNA isolated from heart tissue removed from a Caucasian male fetus who died after 20 weeks gestation from Patau's syndrome (donor A); adrenal gland removed from a 43-year-old Caucasian male (donor B) during nephroureterectomy, regional lymph node excision and unilateral adrenalectomy; kidney cortex removed from a 65-year-old male (donor C) during nephroureterectomy; lung tissue removed from a 14-month-old Caucasian female who died from drowning (donor D); and kidney tissue removed from an 8-year-old Caucasian female who died from a motor vehicle accident (donor E). For donor B, pathology for the associated tumor indicated grade 2 (of 4) renal cell carcinoma in the left kidney with invasion into the renal pelvis. Patient presented with hematuria and anemia. Patient history included benign hypertension and obesity. Previous surgeries included adenotonsillectomy and indirect inguinal hernia repair. The patient was not taking any medications. Family history included benign hypertension and atherosclerotic coronary artery disease in the father. For donc
PITUDIR01	PCDNA2.1	This random primed library was constructed using RNA isolated from pituitary gland tissue removed from a 70-year-old female who died from metastatic adenocarcinoma.
PROSNOT06	PSPORT1	Library was constructed using RNA isolated from the diseased prostate tissue of a 57-year-old Caucasian male during radical prostatectomy, removal of both testes and excision of regional lymph nodes. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated adenocarcinoma (Gleason grade 3+3). Patient history included a benign neoplasm of the large bowel and type I diabetes. Family history included a malignant neoplasm of the prostate and type I diabetes.
SINTFET03	pINCY	Library was constructed using RNA isolated from small intestine tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation.
SMCBUNT01	pINCY	Library was constructed using RNA isolated from untreated bronchial smooth muscle cell tissue removed from a 21-year-old Caucasian male.
THP1AZT01	pINCY	Library was constructed using RNA isolated from THP-1 promonocyte cells treated for three days with 0.8 micromolar 5-aza-2'-deoxycytidine. THP-1 (ATCC TIB 202) is a human promonocyte line derived from peripheral blood of a 1-year-old Caucasian male with acute monocytic leukemia (Int. J. Cancer (1980) 26:171).

Table 6

Library	Vector	Library Description
THYMNOT05	pINCY	Library was constructed using RNA isolated from thymus tissue removed from a 3-year-old Hispanic male during a thymectomy and closure of a patent ductus arteriosus. The patient presented with severe pulmonary stenosis and cyanosis. Patient history included a cardiac catheterization and echocardiogram. Previous surgeries included Blalock-Taussig shunt and pulmonary valvotomy. The patient was not taking any medications. Family history included benign hypertension, osteoarthritis, depressive disorder, and extrinsic asthma in the grandparent(s).
UTREDIT07	pINCY	Library was constructed using RNA isolated from diseased endometrial tissue removed from a female during endometrial biopsy. Pathology indicated in phase endometrium with missing beta 3, Type II defects.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less; Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.06E-6; Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less; Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less

Table 7

Program	Description	Reference	Parameter Threshold
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM hits: Probability value= 1.0E-3 or less; Signal peptide hits: Score= 0 or greater
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	

Table 7

Program	Description	Reference	Parameter Threshold
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. On Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence (AAAI) Press, Menlo Park, CA, and MIT Press, Cambridge, MA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17,

c) a polypeptide comprising a naturally occurring amino acid sequence at least 94% identical to the amino acid sequence of SEQ ID NO:18,

d) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and

e) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36,

c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b), and

e) an RNA equivalent of a)-d).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

19. A method for treating a disease or condition associated with decreased expression of functional NZMS, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional NZMS, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional NZMS, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

a) treating a biological sample containing nucleic acids with the test compound,

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c) quantifying the amount of hybridization complex, and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of NZMS in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of NZMS in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of NZMS in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibodies from said animal, and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of

SEQ ID NO:1-18.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18 in a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18 from a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

a) labeling the polynucleotides of the sample,

b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and

c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.

74. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:19.

75. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:20.

76. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:21.

77. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:22.

78. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:23.

79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:24.

80. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:25.

81. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:26.

82. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:27.

83. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:28.

84. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:29.

85. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:30.

86. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:31.

87. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:32.

88. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:33.

89. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:34.

90. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:35.

91. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:36.

<110> INCYTE GENOMICS, INC.
TANG, Y. Tom
GRIFFIN, Jennifer A.
YUE, Henry
LEE, Ernestine A.
BAUGHN, Mariah R.
DUGGAN, Brendan M.
WALIA, Narinder K.
LEE, Sally
RAMKUMAR, Jayalaxmi
WARREN, Bridget A.
GANDHI, Ameena R.
LU, Dyung Aina M.
LU, Yan
YAO, Monique G.
DING, Li
TRIBOULEY, Catherine M.
SANJANWALA, Madhu M.
ARVIZU, Chandra
HILLMAN, Jennifer L.

<120> ENZYMES

<130> PI-0316 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/251,824; 60/254,312; 60/255,773; 60/256,188;; 60/257,488;
60/262,839; 60/264,402; 60/255,940
<151> 2000-12-07; 2000-12-08; 2000-12-14; 2000-12-15; 2000-12-21;
2001-01-19; 2001-01-26; 2000-12-15

<160> 36
<170> PERL Program

<210> 1
<211> 262
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 8159895CD1

<400> 1

Met	Ser	Arg	Leu	Ser	Trp	Gly	Tyr	Arg	Glu	His	Asn	Gly	Pro	Ile	
1				5				10						15	
His	Trp	Lys	Glu	Phe	Phe	Pro	Ile	Ala	Asp	Gly	Asp	Gln	Gln	Ser	
				20				25						30	
Pro	Ile	Glu	Ile	Lys	Thr	Lys	Glu	Val	Lys	Tyr	Asp	Ser	Ser	Leu	
				35				40						45	
Arg	Pro	Leu	Ser	Ile	Lys	Tyr	Asp	Pro	Ser	Ser	Ala	Lys	Ile	Ile	
				50				55						60	
Ser	Asn	Ser	Gly	His	Ser	Phe	Asn	Val	Asp	Phe	Asp	Asp	Thr	Glu	
				65				70						75	

```
                 80                  85                  90
Leu Arg Gln Val His Leu His Trp Gly Ser Ala Asp Asp His Gly
                 95                 100                 105
Ser Glu His Ile Val Asp Gly Val Ser Tyr Ala Ala Glu Leu His
                110                 115                 120
Val Val His Trp Asn Ser Asp Lys Tyr Pro Ser Phe Val Glu Ala
                125                 130                 135
Ala His Glu Pro Asp Gly Leu Ala Val Leu Gly Val Phe Leu Gln
                140                 145                 150
Ile Gly Glu Pro Asn Ser Gln Leu Gln Lys Ile Thr Asp Thr Leu
                155                 160                 165
Asp Ser Ile Lys Glu Lys Gly Lys Gln Thr Arg Phe Thr Asn Phe
                170                 175                 180
Asp Leu Leu Ser Leu Leu Pro Pro Ser Trp Asp Tyr Trp Thr Tyr
                185                 190                 195
Pro Gly Ser Leu Thr Val Pro Pro Leu Leu Glu Ser Val Thr Trp
                200                 205                 210
Ile Val Leu Lys Gln Pro Ile Asn Ile Ser Ser Gln Gln Leu Ala
                215                 220                 225
Lys Phe Arg Ser Leu Leu Cys Thr Ala Glu Gly Glu Ala Ala Ala
                230                 235                 240
Phe Leu Val Ser Asn His Arg Pro Pro Gln Pro Leu Lys Gly Arg
                245                 250                 255
Lys Val Arg Ala Ser Phe His
                260
```

<210> 2
<211> 502
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2497773CD1

<400> 2

```
Met Ile Pro Ser Lys Lys Asn Ala Val Leu Val Asp Gly Val Val
  1               5                  10                  15
Leu Asn Gly Pro Thr Thr Asp Ala Lys Ala Gly Glu Lys Phe Val
                 20                  25                  30
Glu Glu Ala Cys Arg Leu Ile Met Glu Glu Val Val Leu Lys Ala
                 35                  40                  45
Thr Asp Val Asn Glu Lys Val Cys Glu Trp Arg Pro Pro Glu Gln
                 50                  55                  60
Leu Lys Gln Leu Leu Asp Leu Glu Met Arg Asp Ser Gly Glu Pro
                 65                  70                  75
Pro His Lys Leu Leu Glu Leu Cys Arg Asp Val Ile His Tyr Ser
                 80                  85                  90
Val Lys Thr Asn His Pro Arg Phe Phe Asn Gln Leu Tyr Ala Gly
                 95                 100                 105
Leu Asp Tyr Tyr Ser Leu Val Ala Arg Phe Met Thr Glu Ala Leu
                110                 115                 120
Asn Pro Ser Val Tyr Thr Tyr Glu Val Ser Pro Val Phe Leu Leu
                125                 130                 135
Val Glu Glu Ala Val Leu Lys Lys Met Ile Glu Phe Ile Gly Trp
                140                 145                 150
```

```
Lys Glu Gly Asp Gly Ile Phe Asn Pro Gly Gly Ser Val Ser Asn
                155                 160                 165
Met Tyr Ala Met Asn Leu Ala Arg Tyr Lys Tyr Cys Pro Asp Ile
                170                 175                 180
Lys Glu Lys Gly Leu Ser Gly Ser Pro Arg Leu Ile Leu Phe Thr
                185                 190                 195
Ser Ala Glu Cys His Tyr Ser Met Lys Lys Ala Ala Ser Phe Leu
                200                 205                 210
Gly Ile Gly Thr Glu Asn Val Cys Phe Val Glu Thr Asp Gly Arg
                215                 220                 225
Gly Lys Met Ile Pro Glu Glu Leu Glu Lys Gln Val Trp Gln Ala
                230                 235                 240
Arg Lys Glu Gly Ala Ala Pro Phe Leu Val Cys Ala Thr Ser Gly
                245                 250                 255
Thr Thr Val Leu Gly Ala Phe Asp Pro Leu Asp Glu Ile Ala Asp
                260                 265                 270
Ile Cys Glu Arg His Ser Leu Trp Leu His Val Asp Ala Ser Trp
                275                 280                 285
Gly Gly Ser Ala Leu Met Ser Arg Lys His Arg Lys Leu Leu His
                290                 295                 300
Gly Ile His Arg Ala Asp Ser Val Ala Trp Asn Pro His Lys Met
                305                 310                 315
Leu Met Ala Gly Ile Gln Cys Cys Ala Leu Leu Val Lys Asp Lys
                320                 325                 330
Ser Asp Leu Leu Lys Lys Cys Tyr Ser Ala Lys Ala Ser Tyr Leu
                335                 340                 345
Phe Gln Gln Asp Lys Phe Tyr Asp Val Ser Tyr Asp Thr Gly Asp
                350                 355                 360
Lys Ser Ile Gln Cys Ser Arg Arg Pro Asp Ala Phe Lys Phe Trp
                365                 370                 375
Met Thr Trp Lys Ala Leu Gly Thr Leu Gly Leu Glu Glu Arg Val
                380                 385                 390
Asn Arg Ala Leu Ala Leu Ser Arg Tyr Leu Val Asp Glu Ile Lys
                395                 400                 405
Lys Arg Glu Gly Phe Lys Leu Leu Met Glu Pro Glu Tyr Ala Asn
                410                 415                 420
Ile Cys Phe Trp Tyr Ile Pro Pro Ser Leu Arg Glu Met Glu Glu
                425                 430                 435
Gly Pro Glu Phe Trp Ala Lys Leu Asn Leu Val Ala Pro Ala Ile
                440                 445                 450
Lys Glu Arg Met Met Lys Lys Gly Ser Leu Met Leu Gly Tyr Gln
                455                 460                 465
Pro His Arg Gly Lys Val Asn Phe Phe Arg Gln Val Val Ile Ser
                470                 475                 480
Pro Gln Val Ser Arg Glu Asp Met Asp Phe Leu Leu Asp Glu Ile
                485                 490                 495
Asp Leu Leu Gly Lys Asp Met
                500

<210> 3
<211> 281
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
```

<223> Incyte ID No: 354561CD1

<400> 3
Met Leu Gly Pro Gln Val Trp Ser Ser Val Arg Gln Gly Leu Ser
1 5 10 15
Arg Ser Leu Ser Arg Asn Val Gly Val Trp Ala Ser Gly Glu Gly
20 25 30
Lys Lys Val Asp Ile Ala Gly Ile Tyr Pro Pro Val Thr Thr Pro
35 40 45
Phe Thr Ala Thr Ala Glu Val Asp Tyr Gly Lys Leu Glu Glu Asn
50 55 60
Leu His Lys Leu Gly Thr Phe Pro Phe Arg Gly Phe Val Val Gln
65 70 75
Gly Ser Asn Gly Glu Phe Pro Phe Leu Thr Ser Ser Glu Arg Leu
80 85 90
Glu Val Val Ser Arg Val Arg Gln Ala Met Pro Lys Asn Arg Leu
95 100 105
Leu Leu Ala Gly Ser Gly Cys Glu Ser Thr Gln Ala Thr Val Glu
110 115 120
Met Thr Val Ser Met Ala Gln Val Gly Ala Asp Ala Ala Met Val
125 130 135
Val Thr Pro Cys Tyr Tyr Arg Gly Arg Met Ser Ser Ala Ala Leu
140 145 150
Ile His His Tyr Thr Lys Val Ala Asp Leu Ser Pro Ile Pro Val
155 160 165
Val Leu Tyr Ser Val Pro Ala Asn Thr Gly Leu Asp Leu Pro Val
170 175 180
Asp Ala Val Val Thr Leu Ser Gln His Pro Asn Ile Val Gly Met
185 190 195
Lys Asp Ser Gly Gly Asp Val Thr Arg Ile Gly Leu Ile Val His
200 205 210
Lys Thr Arg Lys Gln Asp Phe Gln Val Leu Ala Gly Ser Ala Gly
215 220 225
Phe Leu Met Ala Ser Tyr Ala Leu Gly Ala Val Gly Gly Val Cys
230 235 240
Ala Leu Ala Asn Val Leu Gly Ala Gln Val Cys Gln Leu Glu Arg
245 250 255
Leu Cys Cys Thr Gly Gln Trp Glu Asp Ala Gln Lys Leu Gln His
260 265 270
Arg Leu Ile Glu Pro Asn Ala Ala Lys Ile Leu
275 280

<210> 4
<211> 433
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7484682CD1

<400> 4
Met Ser Asp Lys Leu Pro Tyr Lys Val Ala Asp Ile Gly Leu Ala
1 5 10 15
Ala Trp Gly Arg Lys Ala Leu Asp Ile Ala Glu Asn Glu Met Pro
20 25 30

```
Gly Leu Met Arg Met Arg Glu Met Tyr Ser Ala Ser Lys Pro Leu
                35                  40                  45
Lys Gly Ala Arg Ile Ala Gly Cys Leu His Met Thr Val Glu Thr
                50                  55                  60
Ala Val Leu Ile Glu Thr Leu Val Ala Leu Gly Ala Glu Val Arg
                65                  70                  75
Trp Ser Ser Cys Asn Ile Phe Ser Thr Gln Asp His Ala Ala Ala
                80                  85                  90
Ala Ile Ala Lys Ala Gly Ile Pro Val Tyr Ala Trp Lys Gly Glu
                95                 100                 105
Thr Asp Glu Glu Tyr Leu Trp Cys Ile Glu Gln Thr Leu His Phe
               110                 115                 120
Lys Asp Gly Pro Leu Asn Met Ile Leu Asp Asp Gly Gly Asp Leu
               125                 130                 135
Thr Asn Leu Ile His Thr Lys Tyr Pro Gln Leu Leu Ser Gly Ile
               140                 145                 150
Arg Gly Ile Ser Glu Glu Thr Thr Thr Gly Val His Asn Leu Tyr
               155                 160                 165
Lys Met Met Ala Asn Gly Ile Leu Lys Val Pro Ala Ile Asn Val
               170                 175                 180
Asn Asp Ser Val Thr Lys Gln Ser Lys Phe Asp Asn Leu Tyr Gly
               185                 190                 195
Cys Arg Glu Ser Leu Ile Asp Gly Ile Lys Arg Ala Thr Asp Val
               200                 205                 210
Met Ile Ala Gly Lys Val Ala Val Val Ala Gly Tyr Gly Asp Val
               215                 220                 225
Gly Lys Gly Cys Ala Gln Ala Leu Arg Gly Phe Gly Ala Arg Val
               230                 235                 240
Ile Ile Thr Glu Ile Asp Pro Ile Asn Ala Leu Gln Ala Ala Met
               245                 250                 255
Glu Gly Tyr Glu Val Thr Thr Met Asp Glu Ala Cys Lys Glu Gly
               260                 265                 270
Asn Ile Phe Val Thr Thr Thr Gly Cys Val Asp Ile Ile Leu Gly
               275                 280                 285
Arg His Phe Glu Gln Met Lys Asp Asp Ala Ile Val Cys Asn Ile
               290                 295                 300
Gly His Phe Asp Val Glu Ile Asp Val Lys Trp Leu Asn Glu Asn
               305                 310                 315
Ala Val Glu Lys Val Asn Ile Lys Pro Gln Val Asp Arg Tyr Arg
               320                 325                 330
Leu Lys Asn Gly Arg Arg Ile Ile Leu Leu Ala Glu Gly Arg Leu
               335                 340                 345
Val Asn Leu Gly Cys Ala Met Gly His Pro Ser Phe Val Met Ser
               350                 355                 360
Asn Ser Phe Thr Asn Gln Val Met Ala Gln Ile Glu Leu Trp Thr
               365                 370                 375
His Pro Asp Lys Tyr Pro Leu Gly Val His Phe Leu Pro Lys Lys
               380                 385                 390
Leu Asp Glu Ala Val Ala Glu Ala His Leu Gly Lys Leu Asn Val
               395                 400                 405
Lys Leu Thr Lys Leu Thr Glu Lys Gln Ala Gln Tyr Leu Gly Met
               410                 415                 420
Pro Ile Asp Gly Pro Phe Lys Pro Asp His Tyr Arg Tyr
               425                 430
```

<210> 5

<211> 308
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7485253CD1

<400> 5
Met Asn Pro Ile Val Val Val His Gly Gly Gly Ala Gly Pro Ile
1 5 10 15
Ser Lys Asp Arg Lys Glu Arg Val His Gln Gly Met Val Arg Ala
20 25 30
Ala Thr Val Gly Tyr Gly Ile Leu Arg Glu Gly Gly Ser Ala Val
35 40 45
Asp Ala Val Glu Gly Ala Val Val Ala Leu Glu Asp Asp Pro Glu
50 55 60
Phe Asn Ala Gly Cys Gly Ser Val Leu Asn Thr Asn Gly Glu Val
65 70 75
Glu Met Asp Ala Ser Ile Met Asp Gly Lys Asp Leu Ser Ala Gly
80 85 90
Ala Val Ser Ala Val Gln Cys Ile Ala Asn Pro Ile Lys Leu Ala
95 100 105
Arg Leu Val Met Glu Lys Thr Pro His Cys Phe Leu Thr Asp Gln
110 115 120
Gly Ala Ala Gln Phe Ala Ala Ala Met Gly Val Pro Glu Ile Pro
125 130 135
Gly Glu Lys Leu Val Thr Glu Arg Asn Lys Lys Arg Leu Glu Lys
140 145 150
Glu Lys His Glu Lys Gly Ala Gln Lys Thr Asp Cys Gln Lys Asn
155 160 165
Leu Gly Thr Val Gly Ala Val Ala Leu Asp Cys Lys Gly Asn Val
170 175 180
Ala Tyr Ala Thr Ser Thr Gly Gly Ile Val Asn Lys Met Val Gly
185 190 195
Arg Val Gly Asp Ser Pro Cys Leu Gly Ala Gly Gly Tyr Ala Asp
200 205 210
Asn Asp Ile Gly Ala Val Ser Thr Thr Gly His Gly Glu Ser Ile
215 220 225
Leu Lys Val Asn Leu Ala Arg Leu Thr Leu Phe His Ile Glu Gln
230 235 240
Gly Lys Thr Val Glu Glu Ala Ala Asp Leu Ser Leu Gly Tyr Met
245 250 255
Lys Ser Arg Val Lys Gly Leu Gly Gly Leu Ile Val Val Ser Lys
260 265 270
Thr Gly Asp Trp Val Ala Lys Trp Thr Ser Thr Ser Met Pro Trp
275 280 285
Ala Ala Ala Lys Asp Gly Lys Leu His Phe Gly Ile Asp Pro Asp
290 295 300
Asp Thr Thr Ile Thr Asp Leu Pro
305

<210> 6
<211> 421
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2397473CD1

<400> 6

```
Met Ser Ala Thr Leu Ile Leu Glu Pro Pro Gly Arg Cys Cys Trp
  1               5                  10                  15
Asn Glu Pro Val Arg Ile Ala Val Arg Gly Leu Ala Pro Glu Gln
                 20                  25                  30
Arg Val Thr Leu Arg Ala Ser Leu Arg Asp Glu Lys Gly Ala Leu
                 35                  40                  45
Phe Arg Ala His Ala Arg Tyr Cys Ala Asp Ala Arg Gly Glu Leu
                 50                  55                  60
Asp Leu Glu Arg Ala Pro Ala Leu Gly Gly Ser Phe Ala Gly Leu
                 65                  70                  75
Glu Pro Met Gly Leu Leu Trp Ala Leu Glu Pro Glu Lys Pro Phe
                 80                  85                  90
Trp Arg Phe Leu Lys Arg Asp Val Gln Ile Pro Phe Val Val Glu
                 95                 100                 105
Leu Glu Val Leu Asp Gly His Asp Pro Glu Pro Gly Arg Leu Leu
                110                 115                 120
Cys Gln Ala Gln His Glu Arg His Phe Leu Pro Pro Gly Val Arg
                125                 130                 135
Arg Gln Ser Val Arg Ala Gly Arg Val Arg Ala Thr Leu Phe Leu
                140                 145                 150
Pro Pro Gly Pro Gly Pro Phe Pro Gly Ile Ile Asp Ile Phe Gly
                155                 160                 165
Ile Gly Gly Gly Leu Leu Glu Tyr Arg Ala Ser Leu Leu Ala Gly
                170                 175                 180
His Gly Phe Ala Thr Leu Ala Leu Ala Tyr Tyr Asn Phe Glu Asp
                185                 190                 195
Leu Pro Asn Asn Met Asp Asn Ile Ser Leu Glu Tyr Phe Glu Glu
                200                 205                 210
Ala Val Cys Tyr Met Leu Gln His Pro Gln Val Lys Gly Pro Gly
                215                 220                 225
Ile Gly Leu Leu Gly Ile Ser Leu Gly Ala Asp Ile Cys Leu Ser
                230                 235                 240
Met Ala Ser Phe Leu Lys Asn Val Ser Ala Thr Val Ser Ile Asn
                245                 250                 255
Gly Ser Gly Ile Ser Gly Asn Thr Ala Ile Asn Tyr Lys His Ser
                260                 265                 270
Ser Ile Pro Pro Leu Gly Tyr Asp Leu Arg Arg Ile Lys Val Ala
                275                 280                 285
Phe Ser Gly Leu Val Asp Ile Val Asp Ile Arg Asn Ala Leu Val
                290                 295                 300
Gly Gly Tyr Lys Asn Pro Ser Met Ile Pro Ile Glu Lys Ala Gln
                305                 310                 315
Gly Pro Ile Leu Leu Ile Val Gly Gln Asp Asp His Asn Trp Arg
                320                 325                 330
Ser Glu Leu Tyr Ala Gln Thr Val Ser Glu Arg Leu Gln Ala His
                335                 340                 345
Gly Lys Glu Lys Pro Gln Ile Ile Cys Tyr Pro Gly Thr Gly His
                350                 355                 360
Tyr Ile Glu Pro Pro Tyr Phe Pro Leu Cys Pro Ala Ser Leu His
                365                 370                 375
Arg Leu Leu Asn Lys His Val Ile Trp Gly Gly Glu Pro Arg Ala
```

```
                380                 385                 390
His Ser Lys Ala Gln Glu Asp Ala Trp Lys Gln Ile Leu Ala Phe
                395                 400                 405
Phe Cys Lys His Leu Gly Gly Thr Gln Lys Thr Ala Val Pro Lys
                410                 415                 420
Leu
```

<210> 7
<211> 155
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7485243CD1

<400> 7

```
Met Ala Pro Ala Arg Ala Gly Phe Cys Pro Leu Leu Leu Leu Leu
  1               5                  10                  15
Leu Leu Gly Leu Trp Val Ala Glu Ile Pro Val Ser Ala Lys Pro
                20                  25                  30
Lys Gly Met Thr Ser Ser Gln Trp Phe Lys Ile Gln His Met Gln
                35                  40                  45
Pro Ser Pro Gln Ala Cys Asn Ser Ala Met Ser Ile Ile Asn Lys
                50                  55                  60
Tyr Thr Glu Arg Cys Lys Asp Leu Asn Thr Phe Leu His Glu Pro
                65                  70                  75
Phe Ser Ser Val Ala Ala Thr Cys Gln Thr Pro Lys Ile Ala Cys
                80                  85                  90
Lys Asn Gly Asp Lys Asn Cys His Gln Ser His Gly Pro Val Ser
                95                 100                 105
Leu Thr Met Cys Lys Leu Thr Ser Gly Lys Tyr Pro Asn Cys Arg
               110                 115                 120
Tyr Lys Glu Lys His Leu Asn Thr Pro Tyr Ile Val Ala Cys Asp
               125                 130                 135
Pro Pro Gln Gln Gly Asp Pro Gly Tyr Pro Leu Val Pro Val His
               140                 145                 150
Leu Asp Lys Val Val
               155
```

<210> 8
<211> 885
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2199285CD1

<400> 8

```
Met Ser His Pro Asp Tyr Arg Met Asn Leu Arg Pro Leu Gly Thr
  1               5                  10                  15
Pro Arg Gly Val Ser Ala Val Ala Gly Pro His Asp Ile Gly Ala
                20                  25                  30
Ser Pro Gly Asp Lys Lys Ser Lys Asn Arg Ser Thr Arg Gly Lys
```

```
                35                  40                  45
Lys Lys Ser Ile Phe Glu Thr Tyr Met Ser Lys Glu Asp Val Ser
                50                  55                  60
Glu Gly Leu Lys Arg Gly Thr Leu Ile Gln Gly Val Leu Arg Ile
                65                  70                  75
Asn Pro Lys Lys Phe His Glu Ala Phe Ile Pro Ser Pro Asp Gly
                80                  85                  90
Asp Arg Asp Ile Phe Ile Asp Gly Val Val Ala Arg Asn Arg Ala
                95                 100                 105
Leu Asn Gly Asp Leu Val Val Val Lys Leu Leu Pro Glu Glu His
               110                 115                 120
Trp Lys Val Val Lys Pro Glu Ser Asn Asp Lys Glu Thr Glu Ala
               125                 130                 135
Ala Tyr Glu Ser Asp Ile Pro Glu Glu Leu Cys Gly His His Leu
               140                 145                 150
Pro Gln Gln Ser Leu Lys Ser Tyr Asn Asp Ser Pro Asp Val Ile
               155                 160                 165
Val Glu Ala Gln Phe Asp Gly Ser Asp Ser Glu Asp Gly His Gly
               170                 175                 180
Ile Thr Gln Asn Val Leu Val Asp Gly Val Lys Lys Leu Ser Val
               185                 190                 195
Cys Val Ser Glu Lys Gly Arg Glu Asp Gly Asp Ala Pro Val Thr
               200                 205                 210
Lys Asp Glu Thr Thr Cys Ile Ser Gln Asp Thr Arg Ala Leu Ser
               215                 220                 225
Glu Lys Ser Leu Gln Arg Ser Ala Lys Val Val Tyr Ile Leu Glu
               230                 235                 240
Lys Lys His Ser Arg Ala Ala Thr Gly Phe Leu Lys Leu Leu Ala
               245                 250                 255
Asp Lys Asn Ser Glu Leu Phe Arg Lys Tyr Ala Leu Phe Ser Pro
               260                 265                 270
Ser Asp His Arg Val Pro Arg Ile Tyr Val Pro Leu Lys Asp Cys
               275                 280                 285
Pro Gln Asp Phe Val Ala Arg Pro Lys Asp Tyr Ala Asn Thr Leu
               290                 295                 300
Phe Ile Cys Arg Ile Val Asp Trp Lys Glu Asp Cys Asn Phe Ala
               305                 310                 315
Leu Gly Gln Leu Ala Lys Ser Leu Gly Gln Ala Gly Glu Ile Glu
               320                 325                 330
Pro Glu Thr Glu Gly Ile Leu Thr Glu Tyr Gly Val Asp Phe Ser
               335                 340                 345
Asp Phe Ser Ser Glu Val Leu Glu Cys Leu Pro Gln Gly Leu Pro
               350                 355                 360
Trp Thr Ile Pro Pro Glu Glu Phe Ser Lys Arg Arg Asp Leu Arg
               365                 370                 375
Lys Asp Cys Ile Phe Thr Ile Asp Pro Ser Thr Ala Arg Asp Leu
               380                 385                 390
Asp Asp Ala Leu Ser Cys Lys Pro Leu Ala Asp Gly Asn Phe Lys
               395                 400                 405
Val Gly Val His Ile Ala Asp Val Ser Tyr Phe Val Pro Glu Gly
               410                 415                 420
Ser Asp Leu Asp Lys Val Ala Ala Glu Arg Ala Thr Ser Val Tyr
               425                 430                 435
Leu Val Gln Lys Val Val Pro Met Leu Pro Arg Leu Leu Cys Glu
               440                 445                 450
Glu Leu Cys Ser Leu Asn Pro Met Ser Asp Lys Leu Thr Phe Ser
```

```
                455                 460                 465
Val Ile Trp Thr Leu Thr Pro Glu Gly Lys Ile Leu Asp Glu Trp
                470                 475                 480
Phe Gly Arg Thr Ile Ile Arg Ser Cys Thr Lys Leu Ser Tyr Glu
                485                 490                 495
His Ala Gln Ser Met Ile Glu Ser Pro Thr Glu Lys Ile Pro Ala
                500                 505                 510
Lys Glu Leu Pro Pro Ile Ser Pro Glu His Ser Ser Glu Glu Val
                515                 520                 525
His Gln Ala Val Leu Asn Leu His Gly Ile Ala Lys Gln Leu Arg
                530                 535                 540
Gln Gln Arg Phe Val Asp Gly Ala Leu Arg Leu Asp Gln Leu Lys
                545                 550                 555
Leu Ala Phe Thr Leu Asp His Glu Thr Gly Leu Pro Gln Gly Cys
                560                 565                 570
His Ile Tyr Glu Tyr Arg Glu Ser Asn Lys Leu Val Glu Glu Phe
                575                 580                 585
Met Leu Leu Ala Asn Met Ala Val Ala His Lys Ile His Arg Ala
                590                 595                 600
Phe Pro Glu Gln Ala Leu Leu Arg Arg His Pro Pro Pro Gln Thr
                605                 610                 615
Arg Met Leu Ser Asp Leu Val Glu Phe Cys Asp Gln Met Gly Leu
                620                 625                 630
Pro Val Asp Phe Ser Ser Ala Gly Ala Leu Asn Lys Ser Leu Thr
                635                 640                 645
Gln Thr Phe Gly Asp Asp Lys Tyr Ser Leu Ala Arg Lys Glu Val
                650                 655                 660
Leu Thr Asn Met Cys Ser Arg Pro Met Gln Met Ala Leu Tyr Phe
                665                 670                 675
Cys Ser Gly Leu Leu Gln Asp Pro Ala Gln Phe Arg His Tyr Ala
                680                 685                 690
Leu Asn Val Pro Leu Tyr Thr His Phe Thr Ser Pro Ile Arg Arg
                695                 700                 705
Phe Ala Asp Val Leu Val His Arg Leu Leu Ala Ala Ala Leu Gly
                710                 715                 720
Tyr Arg Glu Arg Leu Asp Met Ala Pro Asp Thr Leu Gln Lys Gln
                725                 730                 735
Ala Asp His Cys Asn Asp Arg Arg Met Ala Ser Lys Arg Val Gln
                740                 745                 750
Glu Leu Ser Thr Ser Leu Phe Phe Ala Val Leu Val Lys Glu Ser
                755                 760                 765
Gly Pro Leu Glu Ser Glu Ala Met Val Met Gly Ile Leu Lys Gln
                770                 775                 780
Ala Phe Asp Val Leu Val Leu Arg Tyr Gly Val Gln Lys Arg Ile
                785                 790                 795
Tyr Cys Asn Ala Leu Ala Leu Arg Ser His His Phe Gln Lys Val
                800                 805                 810
Gly Lys Lys Pro Glu Leu Thr Leu Val Trp Glu Pro Glu Asp Met
                815                 820                 825
Glu Gln Glu Pro Ala Gln Gln Val Ile Thr Ile Phe Ser Leu Val
                830                 835                 840
Glu Val Val Leu Gln Ala Glu Ser Thr Ala Leu Lys Tyr Ser Ala
                845                 850                 855
Ile Leu Lys Arg Pro Gly Thr Gln Gly His Leu Gly Pro Glu Lys
                860                 865                 870
Glu Glu Glu Glu Ser Asp Gly Glu Pro Glu Asp Ser Ser Thr Ser
```

875 880 885

<210> 9
<211> 282
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2448021CD1

<400> 9
Met Lys Val Lys Val Ile Pro Val Leu Glu Asp Asn Tyr Met Tyr
1 5 10 15
Leu Val Ile Glu Glu Leu Thr Arg Glu Ala Val Ala Val Asp Val
20 25 30
Ala Val Pro Lys Arg Leu Leu Glu Ile Val Gly Arg Glu Gly Val
35 40 45
Ser Leu Thr Ala Val Leu Thr Thr His His His Trp Asp His Ala
50 55 60
Arg Gly Asn Pro Glu Leu Ala Arg Leu Arg Pro Gly Leu Ala Val
65 70 75
Leu Gly Ala Asp Glu Arg Ile Phe Ser Leu Thr Arg Arg Leu Ala
80 85 90
His Gly Glu Glu Leu Arg Phe Gly Ala Ile His Val Arg Cys Leu
95 100 105
Leu Thr Pro Gly His Thr Ala Gly His Met Ser Tyr Phe Leu Trp
110 115 120
Glu Asp Asp Cys Pro Asp Pro Pro Ala Leu Phe Ser Gly Asp Ala
125 130 135
Leu Ser Val Ala Gly Cys Gly Ser Cys Leu Glu Gly Ser Ala Gln
140 145 150
Gln Met Tyr Gln Ser Leu Ala Glu Leu Gly Thr Leu Pro Pro Glu
155 160 165
Thr Lys Val Phe Cys Gly His Glu His Thr Leu Ser Asn Leu Glu
170 175 180
Phe Ala Gln Lys Val Glu Pro Cys Asn Asp His Val Arg Ala Lys
185 190 195
Leu Ser Trp Ala Lys Lys Arg Asp Glu Asp Asp Val Pro Thr Val
200 205 210
Pro Ser Thr Leu Gly Glu Glu Arg Leu Tyr Asn Pro Phe Leu Arg
215 220 225
Val Ala Glu Glu Pro Val Arg Lys Phe Thr Gly Lys Ala Val Pro
230 235 240
Ala Asp Val Leu Glu Ala Leu Cys Lys Glu Arg Ala Arg Phe Glu
245 250 255
Gln Ala Gly Glu Pro Arg Gln Pro Gln Ala Arg Ala Leu Leu Ala
260 265 270
Leu Gln Trp Gly Leu Leu Ser Ala Ala Pro His Asp
275 280

<210> 10
<211> 576
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3187209CD1

<400> 10

```
Met Leu Ala Met Gly Ala Leu Ala Gly Phe Trp Ile Leu Cys Leu
  1               5                  10                  15
Leu Thr Tyr Gly Tyr Leu Ser Trp Gly Gln Ala Leu Glu Glu Glu
             20                  25                  30
Glu Glu Gly Ala Leu Leu Ala Gln Ala Gly Glu Lys Leu Glu Pro
             35                  40                  45
Ser Thr Thr Ser Thr Ser Gln Pro His Leu Ile Phe Ile Leu Ala
             50                  55                  60
Asp Asp Gln Gly Phe Arg Asp Val Gly Tyr His Gly Ser Glu Ile
             65                  70                  75
Lys Thr Pro Thr Leu Asp Lys Leu Ala Ala Glu Gly Val Lys Leu
             80                  85                  90
Glu Asn Tyr Tyr Val Gln Pro Ile Cys Thr Pro Ser Arg Ser Gln
             95                 100                 105
Phe Ile Thr Gly Lys Tyr Gln Ile His Thr Gly Leu Gln His Ser
            110                 115                 120
Ile Ile Arg Pro Thr Gln Pro Asn Cys Leu Pro Leu Asp Asn Ala
            125                 130                 135
Thr Leu Pro Gln Lys Leu Lys Glu Val Gly Tyr Ser Thr His Met
            140                 145                 150
Val Gly Lys Trp His Leu Gly Phe Tyr Arg Lys Glu Cys Met Pro
            155                 160                 165
Thr Arg Arg Gly Phe Asp Thr Phe Phe Gly Ser Leu Leu Gly Ser
            170                 175                 180
Gly Asp Tyr Tyr Thr His Tyr Lys Cys Asp Ser Pro Gly Met Cys
            185                 190                 195
Gly Tyr Asp Leu Tyr Glu Asn Asp Asn Ala Ala Trp Asp Tyr Asp
            200                 205                 210
Asn Gly Ile Tyr Ser Thr Gln Met Tyr Thr Gln Arg Val Gln Gln
            215                 220                 225
Ile Leu Ala Ser His Asn Pro Thr Lys Pro Ile Phe Leu Tyr Ile
            230                 235                 240
Ala Tyr Gln Ala Val His Ser Pro Leu Gln Ala Pro Gly Arg Tyr
            245                 250                 255
Phe Glu His Tyr Arg Ser Ile Ile Asn Ile Asn Arg Arg Arg Tyr
            260                 265                 270
Ala Ala Met Leu Ser Cys Leu Asp Glu Ala Ile Asn Asn Val Thr
            275                 280                 285
Leu Ala Leu Lys Thr Tyr Gly Phe Tyr Asn Asn Ser Ile Ile Ile
            290                 295                 300
Tyr Ser Ser Asp Asn Gly Gly Gln Pro Thr Ala Gly Gly Ser Asn
            305                 310                 315
Trp Pro Leu Arg Gly Ser Lys Gly Thr Tyr Trp Glu Gly Gly Ile
            320                 325                 330
Arg Ala Val Gly Phe Val His Ser Pro Leu Leu Lys Asn Lys Gly
            335                 340                 345
Thr Val Cys Lys Glu Leu Val His Ile Thr Asp Trp Tyr Pro Thr
            350                 355                 360
Leu Ile Ser Leu Ala Glu Gly Gln Ile Asp Glu Asp Ile Gln Leu
            365                 370                 375
Asp Gly Tyr Asp Ile Trp Glu Thr Ile Ser Glu Gly Leu Arg Ser
```

```
                380                 385                 390
Pro Arg Val Asp Ile Leu His Asn Ile Asp Pro Ile Tyr Thr Lys
                395                 400                 405
Ala Lys Asn Gly Ser Trp Ala Ala Gly Tyr Gly Ile Trp Asn Thr
                410                 415                 420
Ala Ile Gln Ser Ala Ile Arg Val Gln His Trp Lys Leu Leu Thr
                425                 430                 435
Gly Asn Pro Gly Tyr Ser Asp Trp Val Pro Pro Gln Ser Phe Ser
                440                 445                 450
Asn Leu Gly Pro Asn Arg Trp His Asn Glu Arg Ile Thr Leu Ser
                455                 460                 465
Thr Gly Lys Ser Val Trp Leu Phe Asn Ile Thr Ala Asp Pro Tyr
                470                 475                 480
Glu Arg Val Asp Leu Ser Asn Arg Tyr Pro Gly Ile Val Lys Lys
                485                 490                 495
Leu Leu Arg Arg Leu Ser Gln Phe Asn Lys Thr Ala Val Pro Val
                500                 505                 510
Arg Tyr Pro Pro Lys Asp Pro Arg Ser Asn Pro Arg Leu Asn Gly
                515                 520                 525
Gly Val Trp Gly Pro Trp Tyr Lys Glu Glu Thr Lys Lys Lys Lys
                530                 535                 540
Pro Ser Lys Asn Gln Ala Glu Lys Lys Gln Lys Lys Ser Lys Lys
                545                 550                 555
Lys Lys Lys Lys Gln Gln Lys Ala Val Ser Gly Ser Thr Cys His
                560                 565                 570
Ser Gly Val Thr Cys Gly
                575
```

<210> 11
<211> 1009
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4507128CD1

```
<400> 11
Met Gly Gln Leu Cys Trp Leu Pro Leu Leu Ala Pro Leu Leu Leu
  1               5                  10                  15
Leu Arg Pro Pro Gly Val Gln Ser Ala Gly Pro Ile Arg Ala Phe
                 20                  25                  30
Val Val Pro His Ser His Met Asp Val Gly Trp Val Tyr Thr Val
                 35                  40                  45
Gln Glu Ser Met Arg Ala Tyr Ala Ala Asn Val Tyr Thr Ser Val
                 50                  55                  60
Val Glu Glu Leu Ala Arg Gly Gln Gln Arg Arg Phe Ile Ala Val
                 65                  70                  75
Glu Gln Glu Phe Phe Arg Leu Trp Trp Asp Gly Val Ala Ser Asp
                 80                  85                  90
Gln Gln Lys Tyr Gln Val Arg Gln Leu Leu Glu Glu Gly Arg Leu
                 95                 100                 105
Glu Phe Val Ile Gly Gly Gln Val Met His Asp Glu Ala Val Thr
                110                 115                 120
His Leu Asp Asp Gln Ile Leu Gln Leu Thr Glu Gly His Gly Phe
                125                 130                 135
```

```
Leu Tyr Glu Thr Phe Gly Ile Arg Pro Gln Phe Ser Trp His Val
                140                 145                 150
Asp Pro Phe Gly Ala Ser Ala Thr Thr Pro Thr Leu Phe Ala Leu
                155                 160                 165
Ala Gly Phe Asn Ala His Leu Gly Ser Arg Ile Asp Tyr Asp Leu
                170                 175                 180
Lys Ala Ala Met Gln Glu Ala Arg Gly Leu Gln Phe Val Trp Arg
                185                 190                 195
Gly Ser Pro Ser Leu Ser Glu Arg Gln Glu Ile Phe Thr His Ile
                200                 205                 210
Met Asp Gln Tyr Ser Tyr Cys Thr Pro Ser His Ile Pro Phe Ser
                215                 220                 225
Asn Arg Ser Gly Phe Tyr Trp Asn Gly Val Ala Val Phe Pro Lys
                230                 235                 240
Pro Pro Gln Asp Gly Val Tyr Pro Asn Met Ser Glu Pro Val Thr
                245                 250                 255
Pro Ala Asn Ile Asn Leu Tyr Ala Glu Ala Leu Val Ala Asn Val
                260                 265                 270
Lys Gln Arg Ala Ala Trp Phe Arg Thr Pro His Val Leu Trp Pro
                275                 280                 285
Trp Gly Cys Asp Lys Gln Phe Phe Asn Ala Ser Val Gln Phe Ala
                290                 295                 300
Asn Met Asp Pro Leu Leu Asp His Ile Asn Ser His Ala Ala Glu
                305                 310                 315
Leu Gly Val Ser Val Gln Tyr Ala Thr Leu Gly Asp Tyr Phe Arg
                320                 325                 330
Ala Leu His Ala Leu Asn Val Thr Trp Arg Val Arg Asp His His
                335                 340                 345
Asp Phe Leu Pro Tyr Ser Thr Glu Pro Phe Gln Ala Trp Thr Gly
                350                 355                 360
Phe Tyr Thr Ser Arg Ser Ser Leu Lys Gly Leu Ala Arg Arg Ala
                365                 370                 375
Ser Ala Leu Leu Tyr Ala Gly Glu Ser Met Phe Thr Arg Tyr Leu
                380                 385                 390
Trp Pro Ala Pro Arg Gly His Leu Asp Pro Thr Trp Ala Leu Gln
                395                 400                 405
Gln Leu Gln Gln Leu Arg Trp Ala Val Ser Glu Val Gln His His
                410                 415                 420
Asp Ala Ile Thr Gly Thr Glu Ser Pro Lys Val Arg Asp Met Tyr
                425                 430                 435
Ala Thr His Leu Ala Ser Gly Met Leu Gly Met Arg Lys Leu Met
                440                 445                 450
Ala Ser Ile Val Leu Asp Glu Leu Gln Pro Gln Ala Pro Met Ala
                455                 460                 465
Ala Ser Ser Asp Ala Gly Pro Ala Gly His Phe Ala Ser Val Tyr
                470                 475                 480
Asn Pro Leu Ala Trp Thr Val Thr Thr Ile Val Thr Leu Thr Val
                485                 490                 495
Gly Phe Pro Gly Val Arg Val Thr Asp Glu Ala Gly His Pro Val
                500                 505                 510
Pro Ser Gln Ile Gln Asn Ser Thr Glu Thr Pro Ser Ala Tyr Asp
                515                 520                 525
Leu Leu Ile Leu Thr Thr Ile Pro Gly Leu Ser Tyr Arg His Tyr
                530                 535                 540
Asn Ile Arg Pro Thr Ala Gly Ala Gln Glu Gly Thr Gln Glu Pro
                545                 550                 555
```

```
Ala Ala Thr Val Ala Ser Thr Leu Gln Phe Gly Arg Arg Leu Arg
                560                 565                 570
Arg Arg Thr Ser His Ala Gly Arg Tyr Leu Val Pro Val Ala Asn
                575                 580                 585
Asp Cys Tyr Ile Val Leu Leu Asp Gln Asp Thr Asn Leu Met His
                590                 595                 600
Ser Ile Trp Glu Arg Gln Ser Asn Arg Thr Val Arg Val Thr Gln
                605                 610                 615
Glu Phe Leu Glu Tyr His Val Asn Gly Asp Val Lys Gln Gly Pro
                620                 625                 630
Ile Ser Asp Asn Tyr Leu Phe Thr Pro Gly Lys Ala Ala Val Pro
                635                 640                 645
Ala Trp Glu Ala Val Glu Met Glu Ile Val Ala Gly Gln Leu Val
                650                 655                 660
Thr Glu Ile Arg Gln Tyr Phe Tyr Arg Asn Met Thr Ala Gln Asn
                665                 670                 675
Tyr Thr Tyr Ala Ile Arg Ser Arg Leu Thr His Val Pro Gln Gly
                680                 685                 690
His Asp Gly Glu Leu Leu Cys His Arg Ile Glu Gln Glu Tyr Gln
                695                 700                 705
Ala Gly Pro Leu Glu Leu Asn Arg Glu Ala Val Leu Arg Thr Ser
                710                 715                 720
Thr Asn Leu Asn Ser Gln Gln Val Ile Tyr Ser Asp Asn Asn Gly
                725                 730                 735
Tyr Gln Met Gln Arg Arg Pro Tyr Val Ser Tyr Val Asn Asn Ser
                740                 745                 750
Ile Ala Arg Asn Tyr Tyr Pro Met Val Gln Ser Ala Phe Met Glu
                755                 760                 765
Asp Gly Lys Ser Arg Leu Val Leu Leu Ser Glu Arg Ala His Gly
                770                 775                 780
Ile Ser Ser Gln Gly Asn Gly Gln Val Glu Val Met Leu His Arg
                785                 790                 795
Arg Leu Trp Asn Asn Phe Asp Trp Asp Leu Gly Tyr Asn Leu Thr
                800                 805                 810
Leu Asn Asp Thr Ser Val Val His Pro Val Leu Trp Leu Leu Leu
                815                 820                 825
Gly Ser Trp Ser Leu Thr Thr Ala Leu Arg Gln Arg Ser Ala Leu
                830                 835                 840
Ala Leu Gln His Arg Pro Val Val Leu Phe Gly Asp Leu Ala Gly
                845                 850                 855
Thr Ala Pro Lys Leu Pro Gly Pro Gln Gln Gln Glu Ala Val Thr
                860                 865                 870
Leu Pro Pro Asn Leu His Leu Gln Ile Leu Ser Ile Pro Gly Trp
                875                 880                 885
Arg Tyr Ser Ser Asn His Thr Glu His Ser Gln Asn Leu Arg Lys
                890                 895                 900
Gly His Arg Gly Glu Ala Gln Ala Asp Leu Arg Arg Val Leu Leu
                905                 910                 915
Arg Leu Tyr His Leu Tyr Glu Val Gly Glu Asp Pro Val Leu Ser
                920                 925                 930
Gln Pro Val Thr Val Asn Leu Glu Ala Val Leu Gln Ala Leu Gly
                935                 940                 945
Ser Val Val Ala Val Glu Glu Arg Ser Leu Thr Gly Thr Trp Asp
                950                 955                 960
Leu Ser Met Leu His Arg Trp Ser Trp Arg Thr Gly Pro Gly Arg
                965                 970                 975
```

```
His Arg Gly Asp Thr Thr Ser Pro Ser Arg Pro Pro Gly Gly Pro
                980                 985                 990
Ile Ile Thr Val His Pro Lys Glu Ile Arg Thr Phe Phe Ile His
                995                1000                1005
Phe Gln Gln Gln
```

<210> 12
<211> 426
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5519834CD1

<400> 12

```
Met Ala Gly Gly His Ser Leu Leu Leu Glu Asn Ala Gln Gln Val
  1               5                  10                  15
Val Leu Val Cys Ala Arg Gly Glu Arg Phe Leu Ala Arg Asp Ala
                20                  25                  30
Leu Arg Ser Leu Ala Val Leu Glu Gly Ala Ser Leu Val Val Gly
                35                  40                  45
Lys Asp Gly Phe Ile Lys Ala Met Gly Pro Ala Asp Val Ile Gln
                50                  55                  60
Arg Gln Phe Ser Gly Glu Thr Phe Glu Glu Ile Ile Asp Cys Ser
                65                  70                  75
Gly Lys Cys Ile Leu Pro Gly Leu Val Asp Ala His Thr His Pro
                80                  85                  90
Val Trp Ala Gly Glu Arg Val His Glu Phe Ala Met Lys Leu Ala
                95                 100                 105
Gly Ala Thr Tyr Met Glu Ile His Gln Ala Gly Gly Gly Ile His
               110                 115                 120
Phe Thr Val Glu Arg Thr Arg Gln Ala Thr Glu Glu Glu Leu Phe
               125                 130                 135
Arg Ser Leu Gln Gln Arg Leu Gln Cys Met Met Arg Ala Gly Thr
               140                 145                 150
Thr Leu Val Glu Cys Lys Ser Gly Tyr Gly Leu Asp Leu Glu Thr
               155                 160                 165
Glu Leu Lys Met Leu Arg Val Ile Glu Arg Ala Arg Arg Glu Leu
               170                 175                 180
Asp Ile Gly Ile Ser Ala Thr Tyr Cys Gly Ala His Ser Val Pro
               185                 190                 195
Lys Gly Lys Thr Ala Thr Glu Ala Ala Asp Asp Ile Ile Asn Asn
               200                 205                 210
His Leu Pro Lys Leu Lys Glu Leu Gly Arg Asn Gly Glu Ile His
               215                 220                 225
Val Asp Asn Ile Asp Val Phe Cys Glu Lys Gly Val Phe Asp Leu
               230                 235                 240
Asp Ser Thr Arg Arg Ile Leu Gln Arg Gly Lys Asp Ile Gly Leu
               245                 250                 255
Gln Ile Asn Phe His Gly Asp Glu Leu His Pro Met Lys Ala Ala
               260                 265                 270
Glu Leu Gly Ala Glu Leu Gly Ala Gln Ala Ile Ser His Leu Glu
               275                 280                 285
Glu Val Ser Asp Glu Gly Ile Val Ala Met Ala Thr Ala Arg Cys
```

```
                290                 295                 300
Ser Ala Ile Leu Leu Pro Thr Thr Ala Tyr Met Leu Arg Leu Lys
                305                 310                 315
Gln Pro Arg Ala Arg Lys Met Leu Asp Glu Gly Val Ile Val Ala
                320                 325                 330
Leu Gly Ser Asp Phe Asn Pro Asn Ala Tyr Cys Phe Ser Met Pro
                335                 340                 345
Met Val Met His Leu Ala Cys Val Asn Met Arg Met Ser Met Pro
                350                 355                 360
Glu Ala Leu Ala Ala Ala Thr Ile Asn Ala Ala Tyr Ala Leu Gly
                365                 370                 375
Lys Ser His Thr His Gly Ser Leu Glu Val Gly Lys Gln Gly Asp
                380                 385                 390
Leu Ile Ile Ile Asn Ser Ser Arg Trp Glu His Leu Ile Tyr Gln
                395                 400                 405
Phe Gly Gly His His Glu Leu Ile Glu Tyr Val Ile Ala Lys Gly
                410                 415                 420
Lys Leu Ile Tyr Lys Thr
                425
```

<210> 13
<211> 583
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2215017CD1

<400> 13

```
Met Asp Ser Leu Lys Gln Glu Asn Lys Asn Asp Arg Ala Lys Lys
  1                 5                  10                  15
Lys Asp Gln Phe Lys Lys Gly Arg Ile Gly Asn Lys Val Gln Thr
                 20                  25                  30
Ile Lys Lys Asn Lys Arg Cys Lys Pro Ser Ser Ala Gly Arg Lys
                 35                  40                  45
Lys Pro Gly Met Tyr Thr Asp Ser Ile Asn Lys Asp Thr Lys Pro
                 50                  55                  60
Pro His Ile Ile Phe Ile Leu Thr Asp Asp Gln Gly Tyr His Asp
                 65                  70                  75
Val Gly Tyr His Gly Ser Asp Ile Glu Thr Pro Thr Leu Asp Arg
                 80                  85                  90
Leu Ala Ala Lys Gly Val Lys Leu Glu Asn Tyr Tyr Ile Gln Pro
                 95                 100                 105
Ile Cys Thr Pro Ser Arg Ser Gln Leu Leu Thr Gly Arg Tyr Gln
                110                 115                 120
Ile His Thr Gly Leu Gln His Ser Ile Ile Arg Pro Gln Gln Pro
                125                 130                 135
Asn Cys Leu Pro Leu Asp Gln Val Thr Leu Pro Gln Lys Leu Gln
                140                 145                 150
Glu Ala Gly Tyr Ser Thr His Met Val Gly Lys Trp His Leu Gly
                155                 160                 165
Phe Tyr Arg Lys Glu Cys Leu Pro Thr Arg Arg Gly Phe Asp Thr
                170                 175                 180
Phe Leu Gly Ser Leu Thr Gly Asn Val Asp Tyr Tyr Thr Tyr Asp
                185                 190                 195
```

```
Asn Cys Asp Gly Pro Gly Val Cys Gly Phe Asp Leu His Glu Gly
                200                 205                 210
Glu Asn Val Ala Trp Gly Leu Ser Gly Gln Tyr Ser Thr Met Leu
                215                 220                 225
Tyr Ala Gln Arg Ala Ser His Ile Leu Ala Ser His Ser Pro Gln
                230                 235                 240
Arg Pro Leu Phe Leu Tyr Val Ala Phe Gln Ala Val His Thr Pro
                245                 250                 255
Leu Gln Ser Pro Arg Glu Tyr Leu Tyr Arg Tyr Arg Thr Met Gly
                260                 265                 270
Asn Val Ala Arg Arg Lys Tyr Ala Ala Met Val Thr Cys Met Asp
                275                 280                 285
Glu Ala Val Arg Asn Ile Thr Trp Ala Leu Lys Arg Tyr Gly Phe
                290                 295                 300
Tyr Asn Asn Ser Val Ile Ile Phe Ser Ser Asp Asn Gly Gly Gln
                305                 310                 315
Thr Phe Ser Gly Gly Ser Asn Trp Pro Leu Arg Gly Arg Lys Gly
                320                 325                 330
Thr Tyr Trp Glu Gly Gly Val Arg Gly Leu Gly Phe Val His Ser
                335                 340                 345
Pro Leu Leu Lys Arg Lys Gln Arg Thr Ser Arg Ala Leu Met His
                350                 355                 360
Ile Thr Asp Trp Tyr Pro Thr Leu Val Gly Leu Ala Gly Gly Thr
                365                 370                 375
Thr Ser Ala Ala Asp Gly Leu Asp Gly Tyr Asp Val Trp Pro Ala
                380                 385                 390
Ile Ser Glu Gly Arg Ala Ser Pro Arg Thr Glu Ile Leu His Asn
                395                 400                 405
Ile Asp Pro Leu Tyr Asn His Ala Gln His Gly Ser Leu Glu Gly
                410                 415                 420
Gly Phe Gly Ile Trp Asn Thr Ala Val Gln Ala Ala Ile Arg Val
                425                 430                 435
Gly Glu Trp Lys Leu Leu Thr Gly Asp Pro Gly Tyr Gly Asp Trp
                440                 445                 450
Ile Pro Pro Gln Thr Leu Ala Thr Phe Pro Gly Ser Trp Trp Asn
                455                 460                 465
Leu Glu Arg Met Ala Ser Val Arg Gln Ala Val Trp Leu Phe Asn
                470                 475                 480
Ile Ser Ala Asp Pro Tyr Glu Arg Glu Asp Leu Ala Gly Gln Arg
                485                 490                 495
Pro Asp Val Val Arg Thr Leu Leu Ala Arg Leu Ala Glu Tyr Asn
                500                 505                 510
Arg Thr Ala Ile Pro Val Arg Tyr Pro Ala Glu Asn Pro Arg Ala
                515                 520                 525
His Pro Asp Phe Asn Gly Gly Ala Trp Gly Pro Trp Ala Ser Asp
                530                 535                 540
Glu Glu Glu Glu Glu Glu Glu Gly Arg Ala Arg Ser Phe Ser Arg
                545                 550                 555
Gly Arg Arg Lys Lys Lys Cys Lys Ile Cys Lys Leu Arg Ser Phe
                560                 565                 570
Phe Arg Lys Leu Asn Thr Arg Leu Met Ser Gln Arg Ile
                575                 580
```

<210> 14
<211> 395
<212> PRT

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7484731CD1

<400> 14

```
Met Val Tyr Pro Gly Ser Ser Leu Ile Arg Gly Asn Asp Glu Ile
  1               5                  10                  15
Pro Arg Ser Pro Pro Val Leu Gln Ala Gly Leu Ser His Arg Val
             20                  25                  30
Leu Ser Gln His Gly Ile Lys Cys Leu Glu Leu Ile Leu Gln Glu
             35                  40                  45
Gln Pro Trp Arg Ser Pro Gln Thr Lys Ala Gln Ser Pro Ser Pro
             50                  55                  60
Val Ser Ile Ile Ser Val Thr Ser Leu Ser Asp Lys Leu Pro Ser
             65                  70                  75
Lys Val Ile Asp Ile Gly Leu Ala Ala Trp Gly Cys Lys Ala Leu
             80                  85                  90
Asn Ile Val Glu Asn Glu Met Pro Gly Leu Met His Met Trp Glu
             95                 100                 105
Leu Tyr Ser Ala Ser Lys Pro Leu Glu Gly Thr His Ser Ala Ser
            110                 115                 120
Cys Leu His Met Thr Met Glu Met Ala Ile His Ile Gly Ser Leu
            125                 130                 135
Ile Thr Leu Gly Ala Pro Ala Ala Thr Ser Ser Ser Pro Trp Thr
            140                 145                 150
Met Cys Trp Leu Pro Leu Pro Arg Leu Ala Phe Gln Phe Thr Pro
            155                 160                 165
Gly Arg Ala Lys Glu Arg Arg Val Pro Met Val His Arg Ala Val
            170                 175                 180
Thr Val Ile Gln Gly Trp Ala Leu Asn Ile Ile Leu Glu Asp Gly
            185                 190                 195
Gly Asp Leu Ala Asn Leu Phe His Thr Lys Met Met Val Asn Gly
            200                 205                 210
Ile Leu Lys Val Pro Ala Ile Asn Val Asn Asp Ser Leu Thr Lys
            215                 220                 225
Ser Glu Phe Asn Lys Leu Tyr Gly Cys Trp Glu Ser Leu Ile Asp
            230                 235                 240
Gly Ile Lys Trp Ala Thr Val Val Met Ile Ala Gly Lys Val Ala
            245                 250                 255
Met Val Ala Gly Tyr Gly Asn Val Gly Lys Gly Cys Ala Gln Ala
            260                 265                 270
Leu Trp Gly Phe Gly Ala His Val Ile Ile Thr Lys Ile Asp Pro
            275                 280                 285
Ile Asn Ala Leu Gln Ala Ala Met Glu Gly Tyr Glu Val Thr Thr
            290                 295                 300
Met Asp Glu Ala Cys Gln Glu Gly Asn Ile Phe Ile Thr Thr Thr
            305                 310                 315
Ala Cys Val Asn Ile Ile Leu Gly Arg His Phe Glu Gln Met Lys
            320                 325                 330
Asp Asp Ala Ile Val Cys Asn Thr Gly His Phe Glu Val Glu Ile
            335                 340                 345
Asn Val Lys Trp Leu Lys Lys Tyr Pro Ile Glu Val His Val Leu
            350                 355                 360
Pro Lys Lys Leu Asp Glu Ala Val Ala Glu Ala His Leu Gly Lys
```

```
                365                 370                 375
Leu Ile Met Lys Leu Thr Lys Leu Thr Glu Lys Gln Ala Gln Tyr
                380                 385                 390
Leu Gly His Leu Pro
                395

<210> 15
<211> 503
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3927361CD1

<400> 15
Met Gln Gln Gln Gln Gln Gln Lys Gln Lys Gln Pro Gly Arg Ser
  1               5                  10                  15
Pro Glu Phe Ser Gly Ala Leu Ala Leu Gln Ala Leu Leu Glu Gly
                20                  25                  30
Gly Thr Ser Arg Arg Ser Ser Ala Ser Ser Arg Ala Arg Arg Pro
                35                  40                  45
Glu Met Pro Pro Asn Leu Ser Ser Ser Pro Cys Arg Ser His Ala
                50                  55                  60
Pro Thr Gly Gly Phe Arg Arg Asn Pro Val Arg Pro Arg Pro Ser
                65                  70                  75
Gly Asn Pro Pro Lys Arg Gly Arg Tyr Leu Val Thr Met Ser Ser
                80                  85                  90
Lys Lys Asn Arg Lys Arg Leu Asn Gln Ser Ala Glu Asn Gly Ser
                95                 100                 105
Ser Leu Pro Ser Ala Ala Ser Ser Cys Ala Glu Ala Arg Ala Pro
               110                 115                 120
Ser Ala Gly Ser Asp Phe Ala Ala Thr Ser Gly Thr Leu Thr Val
               125                 130                 135
Thr Asn Leu Leu Glu Lys Gly Lys Glu Phe Arg Val Tyr Thr Ala
               140                 145                 150
Trp Pro Met Ala Gly Phe Pro Gly Gly Lys Val Gly Leu Ser Glu
               155                 160                 165
Met Ala Gln Lys Asn Val Gly Val Arg Pro Gly Asp Ala Ile Gln
               170                 175                 180
Val Gln Pro Leu Val Gly Ala Val Leu Gln Ala Glu Glu Met Asp
               185                 190                 195
Val Ala Leu Ser Asp Lys Asp Met Glu Ile Asn Glu Glu Glu Leu
               200                 205                 210
Thr Gly Cys Ile Leu Arg Lys Leu Asp Gly Lys Ile Val Leu Pro
               215                 220                 225
Gly Asn Phe Leu Tyr Cys Thr Phe Tyr Gly Arg Pro Tyr Lys Leu
               230                 235                 240
Gln Val Leu Arg Val Lys Gly Ala Asp Gly Met Ile Leu Gly Gly
               245                 250                 255
Pro Gln Ser Asp Ser Asp Thr Asp Ala Gln Arg Met Ala Phe Glu
               260                 265                 270
Gln Ser Ser Met Glu Thr Ser Ser Leu Glu Leu Ser Leu Gln Leu
               275                 280                 285
Ser Gln Leu Asp Leu Glu Asp Thr Gln Ile Pro Thr Ser Arg Ser
               290                 295                 300
```

```
Thr Pro Tyr Lys Pro Ile Asp Asp Arg Ile Thr Asn Lys Ala Ser
                305                 310                 315
Asp Val Cys Trp Met Tyr Thr Glu Pro Trp Arg Trp Gln Trp Thr
                320                 325                 330
Asn Gln Lys Gly Leu Leu Leu Tyr Gly Pro Pro Cys Thr Gly Lys
                335                 340                 345
Thr Met Ile Ala Arg Ala Val Ala Asn Glu Phe Gly Ala Tyr Val
                350                 355                 360
Ser Val Ile Asn Gly Pro Glu Ile Ile Ser Lys His Pro Ser Ile
                365                 370                 375
Ile Phe Ile Asp Glu Leu Asp Ala Leu Cys Pro Lys Arg Glu Gly
                380                 385                 390
Ala Gln Asn Glu Val Glu Lys Arg Val Val Ala Ser Leu Leu Thr
                395                 400                 405
Leu Met Asp Gly Ile Gly Ser Glu Val Ser Glu Gly Gln Val Leu
                410                 415                 420
Val Leu Gly Gly Thr Asn Arg Pro His Ala Leu Asp Ala Ala Leu
                425                 430                 435
Arg Arg Pro Gly Arg Phe Asp Lys Glu Ile Glu Ile Gly Val Pro
                440                 445                 450
Asn Ala Gln Asp Arg Leu Asp Ile Leu Gln Lys Leu Leu Arg Val
                455                 460                 465
Pro His Leu Leu Thr Glu Ala Glu Leu Leu Gln Leu Ala Asn Ser
                470                 475                 480
Ala His Gly Tyr Val Gly Ala Asp Leu Lys Val Leu Cys Asn Glu
                485                 490                 495
Ala Gly Glu Cys Gly Leu Leu Trp
                500
```

<210> 16
<211> 165
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6542758CD1

<400> 16

```
Met Glu Gly Glu Gly Arg Lys Cys Pro Trp Lys Gly Leu Arg Ala
  1               5                  10                  15
Arg Thr Gly Met Gly Gln Glu Val His Gly Ser Cys Trp Ala Leu
                20                  25                  30
Gly Ala Gly Gly Gly Gln Arg Gln Trp Val Gly Arg Ser Met Pro
                35                  40                  45
Pro Leu Ala Pro Gln Leu Cys Arg Ala Val Phe Leu Val Pro Ile
                50                  55                  60
Leu Leu Leu Leu Gln Val Lys Pro Leu Asn Gly Ser Pro Gly Pro
                65                  70                  75
Lys Asp Gly Ser Gln Thr Glu Lys Thr Pro Ser Ala Asp Gln Asn
                80                  85                  90
Gln Glu Gln Phe Glu Glu His Phe Val Ala Ser Ser Val Gly Glu
                95                 100                 105
Met Trp Gln Val Val Asp Met Ala Gln Gln Glu Glu Asp Gln Ser
                110                 115                 120
Ser Lys Thr Ala Ala Val His Lys His Ser Phe His Leu Ser Phe
```

```
                125                 130                 135
Cys Phe Ser Leu Ala Ser Val Met Val Phe Ser Gly Gly Pro Leu
                140                 145                 150
Arg Arg Thr Phe Pro Asn Ile Gln Leu Cys Phe Met Leu Thr His
                155                 160                 165


<210> 17
<211> 453
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3188878CD1

<400> 17
Met Glu Asp Tyr Leu Gln Gly Cys Arg Ala Ala Leu Gln Glu Ser
  1               5                  10                  15
Arg Pro Leu His Val Val Leu Gly Asn Glu Ala Cys Asp Leu Asp
                20                  25                  30
Ser Thr Val Ser Ala Leu Ala Leu Ala Phe Tyr Leu Ala Lys Thr
                35                  40                  45
Thr Glu Ala Glu Glu Val Phe Val Pro Val Leu Asn Ile Lys Arg
                50                  55                  60
Ser Glu Leu Pro Leu Arg Gly Asp Ile Val Phe Phe Leu Gln Lys
                65                  70                  75
Val His Ile Pro Glu Ser Ile Leu Ile Phe Arg Asp Glu Ile Asp
                80                  85                  90
Leu His Ala Leu Tyr Gln Ala Gly Gln Leu Thr Leu Ile Leu Val
                95                 100                 105
Asp His His Ile Leu Ser Lys Ser Asp Thr Ala Leu Glu Glu Ala
                110                 115                 120
Val Ala Glu Val Leu Asp His Arg Pro Ile Glu Pro Lys His Cys
                125                 130                 135
Pro Pro Cys His Val Ser Val Glu Leu Val Gly Ser Cys Ala Thr
                140                 145                 150
Leu Val Thr Glu Arg Ile Leu Gln Gly Ala Pro Glu Ile Leu Asp
                155                 160                 165
Arg Gln Thr Ala Ala Leu Leu His Gly Thr Ile Ile Leu Asp Cys
                170                 175                 180
Val Asn Met Asp Leu Lys Ile Gly Lys Ala Thr Pro Lys Asp Ser
                185                 190                 195
Lys Tyr Val Glu Lys Leu Glu Ala Leu Phe Pro Asp Leu Pro Lys
                200                 205                 210
Arg Asn Asp Ile Phe Asp Ser Leu Gln Lys Ala Lys Phe Asp Val
                215                 220                 225
Ser Gly Leu Thr Thr Glu Gln Met Leu Arg Lys Asp Gln Lys Thr
                230                 235                 240
Ile Tyr Arg Gln Gly Val Lys Val Ala Ile Ser Ala Ile Tyr Met
                245                 250                 255
Asp Leu Glu Ala Phe Leu Gln Arg Ser Asn Leu Leu Ala Asp Leu
                260                 265                 270
His Ala Phe Cys Gln Ala His Ser Tyr Asp Val Leu Val Ala Met
                275                 280                 285
Thr Ile Phe Phe Asn Thr His Asn Glu Pro Val Arg Gln Leu Ala
```

```
                290                 295                 300
Ile Phe Cys Pro His Val Ala Leu Gln Thr Thr Ile Cys Glu Val
                305                 310                 315
Leu Glu Arg Ser His Ser Pro Pro Leu Lys Leu Thr Pro Ala Ser
                320                 325                 330
Ser Thr His Pro Asn Leu His Ala Tyr Leu Gln Gly Asn Thr Gln
                335                 340                 345
Val Ser Arg Lys Lys Leu Leu Pro Leu Leu Gln Glu Ala Leu Ser
                350                 355                 360
Ala Tyr Phe Asp Ser Met Lys Ile Pro Ser Gly Gln Pro Glu Thr
                365                 370                 375
Ala Asp Val Ser Arg Glu Gln Val Asp Lys Glu Leu Asp Arg Ala
                380                 385                 390
Ser Asn Ser Leu Ile Ser Gly Leu Ser Gln Asp Glu Glu Asp Pro
                395                 400                 405
Pro Leu Pro Pro Thr Pro Met Asn Ser Leu Val Asp Glu Cys Pro
                410                 415                 420
Leu Asp Gln Gly Leu Pro Lys Leu Ser Ala Glu Ala Val Phe Glu
                425                 430                 435
Lys Cys Ser Gln Ile Ser Leu Ser Gln Ser Thr Thr Ala Ser Leu
                440                 445                 450
Ser Lys Lys
```

<210> 18
<211> 400
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7500488CD1

<400> 18

```
Met Glu Asp Tyr Leu Gln Gly Cys Arg Ala Ala Leu Gln Glu Ser
  1                 5                  10                  15
Arg Pro Leu His Val Val Leu Gly Asn Glu Ala Cys Asp Leu Asp
                 20                  25                  30
Ser Thr Val Ser Ala Leu Ala Leu Ala Phe Tyr Leu Ala Lys Thr
                 35                  40                  45
Thr Glu Ala Glu Glu Val Phe Val Pro Val Leu Asn Ile Lys Arg
                 50                  55                  60
Ser Glu Leu Pro Leu Arg Gly Asp Ile Val Phe Phe Leu Gln Lys
                 65                  70                  75
Val His Ile Pro Glu Ser Ile Leu Ile Phe Arg Asp Glu Ile Asp
                 80                  85                  90
Leu His Ala Leu Tyr Gln Ala Gly Gln Leu Thr Leu Ile Leu Val
                 95                 100                 105
Asp His His Ile Leu Ser Lys Ser Asp Thr Ala Leu Glu Glu Ala
                110                 115                 120
Val Ala Glu Val Leu Asp His Arg Pro Ile Glu Pro Lys His Cys
                125                 130                 135
Pro Pro Cys His Val Ser Val Glu Leu Val Gly Ser Cys Ala Thr
                140                 145                 150
Leu Val Thr Glu Arg Ile Leu Gln Gly Ala Pro Glu Ile Leu Asp
                155                 160                 165
```

```
Arg Gln Thr Ala Ala Leu Leu His Gly Thr Ile Ile Leu Asp Cys
                170                 175                 180
Val Asn Met Asp Leu Lys Ile Gly Lys Ala Thr Pro Lys Asp Ser
                185                 190                 195
Lys Tyr Val Glu Lys Leu Glu Ala Leu Phe Pro Asp Leu Pro Lys
                200                 205                 210
Arg Asn Asp Ile Phe Asp Ser Leu Gln Lys Ala Lys Phe Asp Val
                215                 220                 225
Ser Gly Leu Thr Thr Glu Gln Met Leu Arg Lys Asp Gln Lys Thr
                230                 235                 240
Ile Tyr Arg Gln Gly Val Lys Val Ala Ile Ser Ala Ile Tyr Met
                245                 250                 255
Asp Leu Glu Ile Cys Glu Val Leu Glu Arg Ser His Ser Pro Pro
                260                 265                 270
Leu Lys Leu Thr Pro Ala Ser Ser Thr His Pro Asn Leu His Ala
                275                 280                 285
Tyr Leu Gln Gly Asn Thr Gln Val Ser Arg Lys Lys Leu Leu Pro
                290                 295                 300
Leu Leu Gln Glu Ala Leu Ser Ala Tyr Phe Asp Ser Met Lys Ile
                305                 310                 315
Pro Ser Gly Gln Pro Glu Thr Ala Asp Val Ser Arg Glu Gln Val
                320                 325                 330
Asp Lys Glu Leu Asp Arg Ala Ser Asn Ser Leu Ile Ser Gly Leu
                335                 340                 345
Ser Gln Asp Glu Glu Asp Pro Pro Leu Pro Pro Thr Pro Met Asn
                350                 355                 360
Ser Leu Val Asp Glu Cys Pro Leu Asp Gln Gly Leu Pro Lys Leu
                365                 370                 375
Ser Ala Glu Ala Val Phe Glu Lys Cys Ser Gln Ile Ser Leu Ser
                380                 385                 390
Gln Ser Thr Thr Ala Ser Leu Ser Lys Lys
                395                 400
```

<210> 19
<211> 1023
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 8159895CB1

<400> 19

```
ttccaccccg agggaccatg tcgaggctca gctggggata ccgcgagcac aacggtccta 60
ttcactggaa ggaatttttc cctattgctg atggtgatca gcaatctcca attgagatta 120
aaaccaaaga agtgaaatat gactcttccc tccgaccact tagtatcaag tatgacccaa 180
gctcagctaa aatcatcagc aacagcggcc attccttcaa tgttgacttt gatgacacag 240
agaacaaatc agttctgcgt ggtggtcctc tcactggaag ctacaggtta cggcaggttc 300
accttcactg ggggtccgct gatgaccacg gctccgagca catagtagat ggagtgagct 360
atgctgcaga gctccatgtt gttcactgga attcagacaa ataccccagc tttgttgagg 420
cagctcatga accagatgga ctggctgtct tgggagtgtt tttacagatt ggtgaaccta 480
attcccaact gcaaaagatt actgacactt tggattccat taaagaaaag ggtaaacaaa 540
ctcgattcac aaattttgac ctattgtctc tgcttccacc atcctgggac tactggacat 600
atcctggttc tcttacagtt ccacctcttc ttgagagtgt cacatggatt gttttaaagc 660
aacctataaa catcagctct caacagctgg ccaaatttcg cagtctcctg tgcacagcgg 720
agggtgaagc agcagctttt ctggtgagca atcaccgccc accacagcct ctaaagggcc 780
```

gcaaagtgag agcctctttc cattaaaaat tgtcaccaat gaactccccc aaacatggct 840
gtggagagac aacaaaacaa aacaaagcac aaaagtctct gccaacaact cttttgtgga 900
attctaattt ataggaaaca ttttagtatg agcttcagtg tcacaaagaa aaccagatct 960
ctctctcttt ttttttatt ttttttagtg atagagtctc actctgtcac ccaggctacg 1020
cgg 1023

<210> 20
<211> 1848
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2497773CB1

<400> 20
cagacaactt agccccagcc actgtctgcc tgcaactgca caagagagct gagtgaaaac 60
aatcttgctg agttcagtca agctccagaa atgtacgggg agatattgat caacaagaga 120
tgattccaag taagaagaat gctgttcttg tggatggggt tgtgctgaat ggtcctacaa 180
cagatgcaaa agctggagaa aaatttgttg aagaggcctg taggctaata atggaagagg 240
tggttttgaa agctacagat gtcaatgaga aggtgtgtga atggaggcct cctgaacaac 300
tgaaacagct tcttgatttg gagatgagag actcaggcga gccaccccat aaactattgg 360
aactctgtcg ggatgtcata cactacagtg tcaaaactaa ccacccaaga tttttcaacc 420
aattgtatgc tggacttgat tattactcct tggtggcccg atttatgacc gaagcattga 480
atccaagtgt ttatacgtat gaggtgtccc cagtgtttct gttagtggaa gaagcggttc 540
tgaagaaaat gattgaattt attggctgga aagaagggga tggaatattt aacccaggtg 600
gctcagtgtc caatatgtat gcaatgaatt tagctagata caaatattgt cctgatatta 660
aggaaaaggg gctgtctggt tcgccaagat taatcctttt cacatctgca gagtgtcatt 720
actctatgaa gaaggcagcc tcttttcttg ggattggcac tgagaatgtt tgctttgtgg 780
aaacagatgg aagaggtaaa atgatacctg aggaactgga gaagcaagtc tggcaagcca 840
gaaaagaggg ggcagcaccg tttcttgtct gtgccacttc tggtacaact gtgttgggag 900
cttttgaccc tctggatgaa atagcagaca tctgcgagag gcacagcctc tggcttcatg 960
tagatgcttc ttggggtggc tcagctttga tgtcgaggaa gcaccgcaag cttctgcatg 1020
gcatccacag ggctgactct gtggcctgga acccacacaa gatgctgatg gctgggatcc 1080
agtgctgtgc tctccttgtg aaagacaaat ctgatcttct taaaaaatgc tactctgcca 1140
aggcatctta cctcttccag caggataaat tctatgatgt gagctatgac acaggagaca 1200
agtctatcca gtgtagcaga agaccagatg cattcaagtt ctggatgacc tggaaggccc 1260
tgggtacatt aggccttgaa gaaagagtta atcgtgctct tgctttatct aggtacctag 1320
tagatgaaat caagaaaaga gaaggattca agttactgat ggaacctgaa tatgccaata 1380
tttgcttttg gtacattcca ccgagcctca gagagatgga agaaggaccc gagttctggg 1440
caaaacttaa tttggtggcc ccagccatta aggagaggat gatgaagaag ggaagcttga 1500
tgctgggcta ccagccgcac cggggaaagg tcaacttctt ccgccaggtg gtgatcagcc 1560
ctcaagtgag ccgggaggac atggacttcc tcctggatga gatagactta ctgggtaaag 1620
acatgtagct gtggctttgg tcccccagag gcatagatcc tatcctggga gagtttagat 1680
ccagaacatc ttggagatac acagtagatt gcagcccttc tgatgagaaa tagggaatac 1740
tcccagtcca ggcccagcaa aaccaaaatg ctaagcaatg aatattaagg actctctagc 1800
tgcctgggca ttactgttgc taaaagaaga aagtttaaaa aaaaaaaa 1848

<210> 21
<211> 1336
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 354561CB1

<400> 21
cctggctggg cttctggctt cagagctggg aattgtggga ggagggccca gatcccactg 60
gagataccga aggagatact ctttagtatt gggatcccag aaaccagtcc tatatctggc 120
cagagggaca ctgggttgtg gcatctctct agctgctacc cagaaggaac agggcccct 180
ggggcctata ggccttgccc ctgaccctgg gaacacccag ctcaggcctg ccccagtggc 240
cacaagtcag ggaggccgca aatttttaac tagaaacatt gatcattaat agggggttag 300
aaagagttca aactaagtct cactctggga catagaccaa ttgtgcttca ggcctcctgc 360
agagatgctg ggtccccaag tctggtcttc tgtgaggcag gggctaagca ggagcttgtc 420
caggaatgtg ggggtctggg cctcagggga ggggaagaag gtggacattg cgggtatcta 480
ccccctgtg accacccct tcactgccac tgcagaggtg gactatggga aactggagga 540
gaatctgcac aaactgggca ccttcccctt ccgaggcttc gtggtccagg gctccaatgg 600
cgagtttcct ttcctgacca gcagtgagcg cctcgaggtg gtgagccgtg tgcgccaggc 660
catgcccaag aacaggctcc tgctagctgg ctccggatgc gagtccactc aagccacagt 720
ggagatgacc gtcagcatgg cccaggtcgg ggctgacgcg gccatggtgg tgacccctg 780
ctactatcgt ggccgcatga gcagtgcggc cctcattcac cactacacca aggttgctga 840
tctctctcca atccctgtgg tgctgtacag tgtcccagcc aacacagggc tggacctgcc 900
tgtggatgca gtggtcacgc tttcccagca cccgaatatt gtgggcatga aggacagcgg 960
tggtgatgtg accaggattg ggctgattgt tcacaagacc aggaagcagg attttcaggt 1020
gttggctgga tcggctggct ttctgatggc cagctatgcc ttgggagctg tgggggcgt 1080
ctgcgccctg gccaatgtcc tgggggctca ggtgtgccag ctggagcgac tgtgctgcac 1140
ggggcaatgg gaagatgccc agaaactgca gcaccgcctc attgagccaa acgctgcgaa 1200
aatcctttga gagaacatcc cagtgtggga actccttgtg actcccagaa aatcctgact 1260
tcggacaaag tcgagcaccc ctgattgtca aacttaagtg tgcattaaaa tttccttggg 1320
ggcgaattaa agttcc 1336

<210> 22
<211> 1302
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7484682CB1

<400> 22
atgtctgata aactgcccta caaagtcgcg gacattggac tggctgcgtg gggacgaaag 60
gccttggaca tagctgaaaa tgagatgcca ggtttgatgc ggatgcggga gatgtactcg 120
gcatccaagc cgctgaaggg tgctcgcatt gctggctgcc tgcacatgac tgtggagacg 180
gctgtcctca ttgagactct cgtggccctg ggtgcggagg tgaggtggtc cagctgcaac 240
atcttctcta ctcaggacca cgcagcagct gccattgcca aggcgggcat tccagtgtat 300
gcctggaagg gcgagacaga cgaagagtac ctgtggtgca ttgagcagac gctgcacttc 360
aaagatggac ctctcaacat gattctggat gatggtggtg acttaactaa cctcatccat 420
accaagtacc cacagcttct gtcaggtatc cgaggcatct ctgaggagac cacgactggg 480
gtccacaacc tctacaagat gatggccaat gggatactga aggtgcctgc catcaatgtc 540
aacgactctg tcaccaagca gagcaagttt gacaacctct atggctgccg ggagtccctc 600
atagatggca tcaaacgagc cacagatgtg atgattgcgg gcaaggtggc ggtggtggca 660
ggctatggcg atgtgggcaa gggctgtgcc caggccctga ggggttttgg ggcccgagtc 720
atcatcaccg agattgaccc catcaatgca ctacaagctg ccatggaggg ttatgaggtg 780
accaccatgg acgaggcctg taaggagggc aatatctttg tcaccaccac aggctgtgtt 840
gatatcattc ttggccggca ctttgaacag atgaaggacg atgccattgt gtgtaacatt 900
ggacactttg acgtggagat tgatgtgaag tggctcaatg agaacgctgt ggagaaggtg 960
aacatcaagc cccaggtgga ccgctacagg ctgaagaatg ggcgccgcat catcctgctg 1020
gctgaaggcc ggctggtcaa cctgggttgt gccatgggcc atcccagctt cgtgatgagc 1080
aactccttca caaaccaggt catggcccag attgagctgt ggacccaccc agataaatac 1140
ccgttggggg ttcacttctt gcctaagaag ctggatgagg cagtggctga agcccacctg 1200
ggcaagctga atgtgaagct gaccaagctg actgagaagc aggcccagta tctgggcatg 1260

cccattgatg gccccttcaa gcctgatcac taccgctact ga 1302

<210> 23
<211> 1428
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7485253CB1

<400> 23
gccggccttc ggggctttat gggaactggg ccgtgcggcg gtcccgccct cgtgcgcagg 60
cgcagaaccg ttgtgaccag agcggtggcg ggctgagcgg tttcgagccg gcgtcgggga 120
gcggcggtac cgggcggctg cggggctggc tcgacccagc tggaggtctc ggcgtccgcg 180
tcctgcggtg ccctgggacc cgccgacatg aatcccatcg tagtggtcca cggcggcgga 240
gccggtccca tctccaagga tcggaaggag cgagtgcacc agggcatggt cagagccgcc 300
accgtgggct acggcatcct ccgggagggc gggagcgccg tggatgccgt agagggagct 360
gtcgtcgccc tggaagacga tcccgagttc aacgcaggtt gtgggtctgt cttgaacaca 420
aatggtgagg ttgaaatgga tgctagtatc atggatggaa aagacctgtc tgcaggagca 480
gtgtccgcag tccagtgtat agcaaatccc attaaacttg ctcggcttgt catggaaaag 540
acacctcatt gctttctgac tgaccaaggc gcagcgcagt ttgcagcagc tatgggggtt 600
ccagagattc ctggagaaaa actggtgaca gagagaaaca aaaagcgcct ggaaaaagag 660
aagcatgaaa aaggtgctca gaaaacagat tgtcaaaaaa acttgggaac cgtgggtgct 720
gttgccttgg actgcaaagg gaatgtagcc tacgcaacct ccacaggcgg tatcgttaat 780
aaaatggtcg gccgcgttgg ggactcaccg tgtctaggag ctggaggtta tgccgacaat 840
gacatcggag ccgtctcaac cacagggcat ggggaaagca tcctgaaggt gaacctggct 900
agactcaccc tgttccacat agaacaagga aagacggtag aagaggctgc ggacctatcg 960
ttgggttata tgaagtcaag ggttaaaggt ttaggtggcc tcatcgtggt tagcaaaaca 1020
ggagactggg tggcaaagtg gacctccacc tccatgccct gggcagccgc caaggacggc 1080
aagctgcact tcggaattga tcctgacgat actactatca ccgaccttcc ctaagccgct 1140
ggaagattgt attccagatg ctagcttaga ggtcaagtac agtctcctca tgagacatag 1200
cctaatcaat tagatctaga attggaaaaa ttgtcccgtc tgtcacttgt tttgttgcct 1260
taataagcat ctgaatgttt ggttgtgggg cgggttctga agcaatgaga gaaatgcccg 1320
tattaggagg attacttgag ccctggaggt caaagctgag gtgagccatg attactccac 1380
tgcactccag cctgggcaac agagccaggc cctgtatcaa aaaaaaaa 1428

<210> 24
<211> 1393
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2397473CB1

<400> 24
atgtcagcaa cgctgatcct ggagccccca ggccgctgct gctggaacga gccggtgcgc 60
attgccgtgc gcggcctggc cccggagcag cgggttacgc tgcgcgcgtc cctgcgcgac 120
gagaagggcg cgctcttccg ggcccacgcg cgctactgcg ccgacgcccg cggcgagctg 180
gacctggagc gcgcacccgc gctgggcggc agcttcgcgg gactcgagcc catggggctg 240
ctctgggccc tggaacccga gaagcctttt tggcgcttcc tgaagcggga cgtacagatt 300
ccttttgtcg tggagttgga ggtgctggac ggccacgacc ccgagcctgg acggctgctg 360
tgccaggcgc agcacgagcg ccacttcctc ccgccagggg tgcggcgcca gtcggtgcga 420
gcgggccggg tgcgcgccac gctcttcctg ccgccaggac ctggaccctt cccagggatc 480
attgacatct ttggtattgg agggggcctc ttggaatatc gagccagcct ccttgctggc 540

catggctttg ccacgttggc tctagcttat tataactttg aagatctccc caataacatg 600
gacaacatat ccctggagta cttcgaagaa gccgtatgct acatgcttca acatccccag 660
gtaaaaggcc caggcattgg gcttttgggc atttctctag gagctgatat ttgtctctca 720
atggcctcat tcttgaagaa tgtctcagcc acagtttcca tcaatggatc tgggatcagt 780
gggaacacag ccatcaacta taagcacagt agcattccac cattgggcta tgacctgagg 840
agaatcaagg tagctttctc aggcctcgtg gacatcgtgg atataaggaa tgctctcgta 900
ggagggtaca agaaccccag catgattcca atagagaagg cccaggggcc catcctgctc 960
attgttggtc aggatgacca taactggaga agtgagttgt atgcccaaac agtctctgaa 1020
cggttacagg cccatggaaa ggaaaaaccc cagatcatct gttaccctgg gactgggcat 1080
tacatcgagc ctccttactt ccccctgtgc ccagcttccc ttcacagatt actgaacaaa 1140
catgttatat ggggtgggga gcccagggct cattctaagg cccaggaaga tgcctggaag 1200
caaattctag ccttcttctg caaacacctg ggaggtaccc agaaaacagc tgtccctaaa 1260
ttgtaatgca tttgtctgtt gttgacatga gagattcaag atcagattct agtgttcagt 1320
aaccctatgt gaatcagatg tctcctggat aacattaaag ccatgtcttt gtcattaaaa 1380
aaaaaaaaaa aaa 1393

<210> 25
<211> 567
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7485243CB1

<400> 25
gcgtggacac cacctcagcc cactgagcag gagtcacagc acgaagacca agcgcaaagc 60
gacccctgcc ctccatcctg actgctcctc ctaagagaga tggcaccggc cagagcagga 120
ttctgccccc ttctgctgct tctgctgctg gggctgtggg tggcagagat cccagtcagt 180
gccaagccca agggcatgac ctcatcacag tggtttaaaa ttcagcacat gcagcccagc 240
cctcaagcat gcaactcagc catgagcatc atcaataagt acacagaacg gtgcaaagac 300
ctcaacacct tcctgcacga gcctttctcc agtgtggccg ccacctgcca gacccccaaa 360
atagcctgca agaatggcga taaaaactgc caccagagcc acgggcccgt gtccctgacc 420
atgtgtaagc tcacctcagg gaagtatccg aactgcaggt acaaagagaa gcacctgaac 480
acaccttaca tagtggcctg tgaccctcca caacagggtg acccagggta cccacttgtt 540
cctgtgcact tggataaagt tgtctaa 567

<210> 26
<211> 3519
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2199285CB1

<400> 26
atggtcgagc tcggcatcac ttgtacggcg cagtgtgctg gaaaggcatg gacaattcca 60
ccagaggagt tcagcaagag aaatcagaga aatccacgcc atactctgtt agtattcctt 120
ctgtttcagg ctcaatttca ccagcctgcc caagactctt agcaactgcg ggactgcggc 180
ggcgccggcc tccggggaga aacgcgaatg acaacagagc tgctcaaggc gggaactctg 240
agctaagcag tggaggtttc tctggatctg gagagaagag tgaccttgga gccaataatg 300
agccatcctg actacagaat gaacctccgg cccctgggga cccccagagg tgtgtctgct 360
gtggctggtc cacatgacat tggtgcttcg ccaggtgaca aaaagtcaaa gaacaggtcc 420
acacgaggga agaaaaagag catatttgaa acttacatgt ccaaggagga tgtttcagaa 480
ggcttgaaga gaggaacact catccagggt gtattgagaa ttaatccaaa gaagtttcat 540

gaagccttca ttccttcccc ggatggtgat cgagacattt ttattgatgg ggttgttgct 600
cgtaatagag ccttaaatgg ggatctggtg gtcgtgaaac tgcttcccga ggagcattgg 660
aaggtagtta aaccagagag caatgacaaa gaaacagaag ctgcgtatga atcagatatc 720
cccgaggagc tctgtggaca ccatctcccg caacagtccc tgaaaagcta taatgacagt 780
cctgatgtca ttgtagaggc tcagtttgat ggcagcgact cagaagatgg acatggcatc 840
acacaaaatg tgctggttga tggtgttaag aaactctcag tttgtgtttc tgagaaagga 900
agagaggatg gtgatgcacc ggttacaaaa gatgagacca cctgcatttc acaagacaca 960
agagctttat cggagaaatc cctgcaaaga tcagcaaagg tggtttacat cttggagaaa 1020
aaacattctc gagcagcaac cggcttcctc aaactcttgg ctgataagaa cagcgaactg 1080
tttaggaaat acgccctgtt ttctccctca gaccaccgag tgcctagaat ttatgtgcct 1140
ctcaaggact gtccccagga ctttgtggca cggcctaaag attatgccaa cacactgttc 1200
atctgccgca ttgtggactg gaaggaggac tgcaattttg ccctggggca gctggctaag 1260
agtcttgggc aggctggtga aattgagcct gaaacagaag gaatactaac agagtatggc 1320
gtggatttct ctgatttctc ttcagaagtt ctagaatgtc ttcctcaagg cctgccatgg 1380
acaattccac cagaggagtt cagcaagaga agggatttaa gaaaagactg tatcttcacc 1440
attgacccat caaccgcccg agacctcgat gatgccctct cctgcaagcc actcgctgac 1500
ggcaacttca aagtgggagt tcacattgct gacgtgagtt actttgttcc ggagggatct 1560
gatctggata aagtggctgc cgagagggct acaagcgtct acttggttca aaaggtggtc 1620
cccatgcttc ccaggctgct gtgtgaggag ctgtgcagcc tcaaccccat gtccgacaag 1680
ctgaccttct ctgtgatctg gacactgact ccagagggca agatccttga tgaatggttt 1740
ggccggacca tcatccgctc ctgcaccaaa cttagctacg agcatgcaca gagcatgatt 1800
gaaagcccaa ctgagaaaat ccctgcgaaa gagctgcccc ccatttcccc agagcatagc 1860
agcgaggagg tacaccaggc cgtcttgaat ctccacggaa ttgccaagca gttacgccag 1920
cagcgctttg tggacggcgc acttcgtttg gatcagctaa agcttgcttt cactctggac 1980
cacgagaccg gattgcctca aggatgtcat atctatgagt accgcgagag caacaagctc 2040
gtggaggagt tcatgctctt ggccaacatg gcagtggccc acaagatcca ccgcgccttc 2100
cccgagcagg ccctgctgcg ccggcacccc ccgccccaaa caaggatgct cagtgacctg 2160
gtggaattct gcgaccagat ggggctgccc gtggacttca gctccgcagg agccctcaat 2220
aaaagcctga cccaaacatt tggagatgac aagtactcac tggcccgcaa ggaggtgctc 2280
accaacatgt gctcccggcc catgcagatg gcactgtact tctgctcggg gctgctgcag 2340
gacccagcgc agttccggca ctacgcgctc aatgtgcccc tgtacacaca cttcacctcg 2400
cccatccgcc gctttgccga cgtcctggtg caccgcctcc tggctgccgc gttaggctat 2460
agggagcgac tagacatggc gcccgatacc ctgcagaaac aggcggacca ctgtaacgac 2520
cgccgcatgg cgtccaagcg cgtgcaggag ctcagtacca gtctcttctt tgctgttctg 2580
gtcaaggaga gtggcccccc ggagtcagaa gccatggtga tgggcatcct gaagcaagcc 2640
ttcgacgtgc tggtgctgcg ctacggcgtg cagaagcgca tctactgcaa cgcactggcc 2700
ctgcggtccc accacttcca gaaggtgggc aagaagccgg aactcacgct ggtctgggag 2760
cctgaggaca tggagcagga gccagcacag caggtcatca ccatcttcag cctggtggag 2820
gtggtcctgc aggcagagtc cacagccctc aagtacagcg ccatcctgaa gcggccaggc 2880
acccagggcc acctgggccc tgagaaggag gaggaggagt ctgacggtga gcccgaggac 2940
tcaagcacca gctgagctcc accagccgcc tgccccgcct gccccgcctg cctgtcccgc 3000
cacactggct ttaggacctg ttgacacgga ggggggtttt taatttggtt tttaacaact 3060
caggggtttg tttttatttt tatttaattt ttgcagctca acttttaaac aaactgcagg 3120
ggagagggtg gggctggaag gaaggctgag gcctggtcag cagtgacccc agcagagcag 3180
gccccagtcc tcctgggagg ctggcccccc ttttttctgg gccctactgc cctcctctgc 3240
ccaggaaatg gggggggttc agcaactcag tgtcacagaa taaaatcaag tgtggagtgc 3300
catctggtgt gtagggcgcc tctgggaagc ctgggcagca gaatgcccct tgcacccagg 3360
gcaagggacc cagttcaggc ttcacccctc gctgctgagc cgatgtcaac acctggaact 3420
ttcctgtcag ttccaacacg attcagagct ggctgcctgg cagatgattg atactggagt 3480
ctcattctgc ctgattaaaa atggaattag tatgcaaaa 3519

<210> 27
<211> 1291
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2448021CB1

<400> 27
ctcagcacag ccgtgccggt gaggcgggcg gcgggggaac gcggctgtcc cggcccttcc 60
tagggtgtgg agagcgggcc ccgccctgaa ggggcaccgt gggctggggg gcctgttttg 120
gagcaggcac cggtggccga gctccgtgac catgaaggtc aaggtcatcc ccgtgctcga 180
ggacaactac atgtacctgg tcatcgagga gctcacgcgc gaggcggtgg ccgtggacgt 240
ggctgtgccc aagaggctgc tggagatcgt gggccgggag ggggtgtctc tgaccgctgt 300
gctgaccacc caccatcact gggaccacgc gcggggaaac ccggagctgg cgcggcttcg 360
tcccgggctg gcggtgctgg gcgcggacga gcgcatcttc tcgctgacgc gcaggctggc 420
gcacggcgag gagctgcggt tcggggccat ccacgtgcgt tgcctcctga cgcccggcca 480
caccgccggc cacatgagct acttcctgtg ggaggacgat tgcccggacc cacccgccct 540
gttctcgggc gacgcgctgt cggtggccgg ctgcggctcg tgcctggagg gcagcgccca 600
gcagatgtac cagagcctgg ccgagctggg taccctgccc cccgagacga aggtgttctg 660
cggccacgag cacacgctta gcaacctgga gtttgcccag aaagtggagc cctgcaacga 720
ccacgtgaga gccaagctgt cctgggctaa gaagagggat gaggatgacg tgcccactgt 780
gccgtcgact ctgggcgagg agcgcctcta caacccctcc ctgcgggtgg cagaggagcc 840
ggtgcgcaag ttcacgggca aggcggtccc cgccgacgtc ctggaggcgc tatgcaagga 900
gcgggcgcgc ttcgaacagg cgggcgagcc gcggcagcca caggcgcggg ccctccttgc 960
gctgcagtgg gggctcctga gtgcagcccc acacgactga gccacccaga ccctcacagg 1020
gctggggcct gcgtccctcc tcgtgacctc ggccagctgg acccacatga gggccacctc 1080
tggaaccttc ttcgaggccc tggccagcca tctgcccagc ctcggagggt gggcaacctg 1140
gtgcttcccg ggtggacaca caggaccact cagtggggcc tgtgtgggcg ccgagacctg 1200
ggtgtctggg aagtggggca cacggggcct ccgaactatg aataaagctt tgaaaggcca 1260
aaaaaaaaaa aaaaaaaaaa aaaaaatggt t 1291

<210> 28
<211> 3072
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3187209CB1

<400> 28
gaagagcgag ccctccttgt tcttccggag tcccatccat taagccatca cttctggaag 60
attaaagttg tcggacatgg tgacagctga gaggagagga ggatttcttg ccaggtggag 120
agtcttcacc gtctgttggg tgcatgtgtg cgcccgcagc ggcgcggggc gcgtggttct 180
ccgcgtggag tctcacctgg gacctgagtg aatggctccc aggggctgtg cggggcatcc 240
gcctccgcct tctccacagg cctgtgtctg tcctggaaag atgctagcaa tgggggcgct 300
ggcaggattc tggatcctct gcctcctcac ttatggttac ctgtcctggg gccaggcctt 360
agaagaggag gaagaagggg ccttactagc tcaagctgga gagaaactag agcccagcac 420
aacttccacc tcccagcccc atctcatttt catcctagcg gatgatcagg gatttagaga 480
tgtgggttac cacggatctg agattaaaac acctactctt gacaagctcg ctgccgaagg 540
agttaaactg gagaactact atgtccagcc tatttgcaca ccatccagga gtcagtttat 600
tactggaaag tatcagatac acaccggact tcaacattct atcataagac ctacccaacc 660
caactgttta cctctggaca atgccaccct acctcagaaa ctgaaggagg ttggatattc 720
aacgcatatg gtcggaaaat ggcacttggg tttttacaga aaagaatgca tgcccaccag 780
aagaggattt gataccttt ttggctccct tttgggaagt ggggattact atacacacta 840
caaatgtgac agtcctggga tgtgtggcta tgacttgtat gaaaacgaca atgctgcctg 900
ggactatgac aatggcatat actccacaca gatgtacact cagagagtac agcaaatctt 960
agcttcccat aaccccacaa agcctatatt tttatatatt gcctatcaag ctgttcattc 1020
accactgcaa gctcctggca ggtatttcga acactaccga tccattatca acataaacag 1080

gaggagatat gctgccatgc tttcctgctt agatgaagca atcaacaacg tgacattggc 1140
tctaaagact tatggtttct ataacaacag cattatcatt tactcttcag ataatggtgg 1200
ccagcctacg gcaggaggga gtaactggcc tctcagaggt agcaaaggaa catattggga 1260
aggagggatc cgggctgtag gctttgtgca tagcccactt ctgaaaaaca agggaacagt 1320
gtgtaaggaa cttgtgcaca tcactgactg gtaccccact ctcatttcac tggctgaagg 1380
acagattgat gaggacattc aactagatgg ctatgatatc tgggagacca taagtgaggg 1440
tcttcgctca ccccgagtag atattttgca taacattgac cccatataca ccaaggcaaa 1500
aaatggctcc tgggcagcag gctatgggat ctggaacact gcaatccagt cagccatcag 1560
agtgcagcac tggaaattgc ttacaggaaa tcctggctac agcgactggg tccccctca 1620
gtctttcagc aacctgggac cgaaccggtg gcacaatgaa cggatcacct tgtcaactgg 1680
caaaagtgta tggcttttca acatcacagc cgacccatat gagagggtgg acctatctaa 1740
caggtatcca ggaatcgtga agaagctcct acggaggctc tcacagttca acaaaactgc 1800
agtgccggtc aggtatcccc ccaaagaccc cagaagtaac cctaggctca atggaggggt 1860
ctggggacca tggtataaag aggaaaccaa gaaaaagaag ccaagcaaaa atcaggctga 1920
gaaaaagcaa aagaaaagca aaaaaaagaa gaagaaacag cagaaagcag tctcaggttc 1980
aacttgccat tcaggtgtta cttgtggata agcacaaata tttcctgttt ggttaaactt 2040
taatcagttc ttatctttca tctgtttcct aggtaaacca gcaaatttgg ctcgataata 2100
tcgctggcct aagcgtcagg cttgttttca tgctgtgcca ctccagagac ttctgccacc 2160
tggccgccac actgaaaact gtcctgctca gtgccaaggt gctactcttg caagccacac 2220
ttagagagag tggagatgtt tatttctctc gctcctttag aaaacgtggt gagtcctgag 2280
ttccactgct gtgcttcagt caactgacca aacactgctt tgaattatag gaggagaaca 2340
ataacctacc atccgcaagc atgctaattt gatggaagtt acagggtagc atgattaaaa 2400
ctacctttga taaattacag tcaaagattg tgtcacctca aaggccttga agaatatatt 2460
ttcttggtga atttttgtat gtctgtcata tgacacttgg gtttttttaat taattctatt 2520
ttatatatat aaatatatgt ttcttttcct gtgaaaagct gtttttctca catgtgaaca 2580
gcttgcacct cattttacca tgcgtgaggg aatggcaaat aagaatgttt gagcacactg 2640
cccacaatga atgtaactat tttctaaaca ctttactaga agaacatttc agtataaaaa 2700
acctaattta tttttacaga aaaatatttt gttgttttta taaaaagtta tgcaaatgac 2760
ttttatttttt atttcctgca taccattaga agaattttat ttcatttctt caaattatca 2820
agcactgtaa tactataaat taatgtaata ctgtgtgaat tcagactata aaaaacatca 2880
ttcagaaaac tttataatcg tcattgttca atcaagattt tgaatgtaat aagatgaata 2940
tattccttac aaattacttg gaaattcaat gtttgtgcag agttgagaca actttattgt 3000
ttctatcata aactatttat gtatcttaat tattaaaatg atttacttta tggcactaga 3060
aaaaaaaaaa aa 3072

<210> 29
<211> 4117
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4507128CB1

<400> 29
ggcaccttcc cggcctgccg cagggatggg gcagctgtgc tggctgccgc tgctggcacc 60
gctcctgttg ctgcgaccgc caggggtcca gtccgccggc cccatccggg ccttcgtggt 120
gccccacagc cacatggacg tgggctgggt ctacactgtg caggaaagca tgcgggcgta 180
cgccgccaat gtctacacct cagtggtgga agagctggcc cgcggccagc agcgccggtt 240
catcgctgtg gagcaggagt ttttccggct gtggtgggat ggcgtcgcct cggaccagca 300
gaaataccag gtccgccagc tcctggagga aggacgcctg gaatttgtca tcggaggcca 360
ggtcatgcat gacgaggctg tgacgcacct tgatgaccag atcctgcagc tcacagaagg 420
acacgggttt ctctatgaaa catttgggat ccggccacag ttctcctggc acgttgaccc 480
gtttggcgcc tctgccacga cgcccaccct atttgcgctg gcgggcttca atgcccacct 540
cggctcccgg atcgactacg acctgaaggc agccatgcag gaggcccggg ggctgcagtt 600
cgtgtggcga gggtccccat ccctctcaga gcggcaggaa atcttcacgc acatcatgga 660

ccagtacagc tactgcaccc cgtcccacat ccctttctcc aacaggtcag gattttactg 720
gaatggcgtg gctgtcttcc ccaagcctcc ccaagatggg gtgtacccca acatgagtga 780
gcctgtcacc ccagccaaca tcaacctcta tgccgaggcc ctggtggcca acgtgaagca 840
gagggccgcc tggttccgga caccgcacgt cctctggccc tggggatgtg acaagcagtt 900
cttcaatgcc tcggtgcagt ttgccaacat ggacccgctg ctggaccaca tcaacagcca 960
tgctgccgag ctcggtgtct cggtgcagta tgccacgctg ggcgactact tccgtgccct 1020
gcacgctctc aatgtcacct ggcgtgtccg cgaccaccac gacttcctgc cctattccac 1080
agaaccattc caggcctgga cgggcttcta cacgtcccgc agctcactga aggggctggc 1140
ccggcgagcc agcgccttgt tgtatgccgg ggagtccatg ttcacacgct acctgtggcc 1200
ggcccccgt gggcatctgg accccacctg ggccctgcag cagctccagc agcttcgctg 1260
ggccgtctcc gaggtccagc accatgatgc catcactggg actgagtccc ccaaggtgag 1320
agacatgtac gcaacgcacc tggcctcggg gatgctgggc atgcgcaagc tgatggcctc 1380
catcgtccta gatgagctcc agccccaggc acccatggcg gccagctccg atgcaggacc 1440
tgcaggacat tttgcctcgg tctacaaccc gctggcctgg acggtcacca ccatcgtcac 1500
cctgactgtt ggtttccctg gagtccgcgt cacagatgag gcgggccacc cagtgccctc 1560
gcagatccag aactcaacag agaccccatc tgcgtatgac ctgcttattc tgaccacaat 1620
cccaggcctc agttaccggc actacaacat cagacccact gcaggggccc aagagggcac 1680
ccaggagccg gctgccactg tggcgagcac ccttcaattt ggccgcaggc tgaggagacg 1740
caccagccat gcgggcaggt acttggtgcc tgtggcaaac gactgctaca ttgtgctgct 1800
cgaccaggat accaacctga tgcacagcat ctgggagaga cagagtaacc gaacggtgcg 1860
cgtgacccag gaattcctgg agtaccacgt caacggggat gtgaaacagg gccccatttc 1920
cgataactac ctgttcacac cgggcaaggc cgcggtgcct gcgtgggaag ctgtggaaat 1980
ggagattgtg gcgggacagc ttgtgactga gatccggcag tacttctaca ggaacatgac 2040
agcacagaat tacacgtatg caatccgctc ccggctcacc catgtgccgc agggccatga 2100
cggggagctg ctctgccacc ggatagagca ggagtaccaa gccggccccc tggagctgaa 2160
ccgtgaggct gtcctgagga ccagcaccaa cctaaacagc cagcaggtca tctactcaga 2220
caacaacggc taccagatgc agcggaggcc ctacgtttcc tatgtgaaca acagcatcgc 2280
ccggaattac taccccatgg ttcagtcggc cttcatggag gatggcaaaa gcaggcttgt 2340
gttgctgtcg gagcgggcac atggcatctc cagccaaggg aatgggcagg tggaggtcat 2400
gctccaccgg cggctgtgga acaacttcga ctgggacctg ggctacaacc tcacgctgaa 2460
cgacacctca gtcgtccacc cagtgctctg gcttctgctg ggatcctggt ccctcaccac 2520
tgccctgcgc cagaggagcg cactggcgct gcagcacagg cccgtggtgc tgttcggaga 2580
cctcgctggg actgcgccga agctcccagg accccagcag caagaggccg tgacgctgcc 2640
cccgaatctt cacctgcaga tcctgagcat ccctggctgg cgctacagct ccaaccacac 2700
ggagcactct cagaatctcc ggaaaggcca tcgaggggaa gcccaggctg acctccgccg 2760
tgtcctgctg cggctctacc acctgtatga agtgggcgag gacccagtcc tgtctcagcc 2820
agtgacagtg aatctggagg ctgtgctgca ggcgctgggg tccgtggtgg cagtggagga 2880
gcgctcgctc acagggacct gggatttgag catgctgcac cgctggagct ggaggacggg 2940
gcctggccgc cacagaggtg acaccacctc tccctcgagg ccaccaggag gccccatcat 3000
caccgtccac ccaaaggaaa tccggacgtt ctttattcac tttcaacagc agtgagccct 3060
gggcagatgc cccgtcccca gggcttcccc caggaactcc atgtaacaga acagacccag 3120
gacagggaaa agcagtgcgg agggatggga ctggggagtc agctgctcat ctgcaggcta 3180
atggcaggaa atggtcatat ttgggggtttt tccctaattt ttttaaacaa aaattacatt 3240
acaagatcca ggttcttccc ccccacactc aatcaagcca gccctctcct cttctgtcac 3300
gtaaaggata tttggcacac tcatgcgtca ttcattcaca aaacacaaac ccaggacttt 3360
ctgcctaagg cagaacacaa gactcacagc agcaccgaag cgcatctgcc gtccgggccc 3420
tgccaggctt gccaggctgc cagtggtaac tgtggaccta ctgcgtgcca cgtgttttca 3480
tagactcatc ccatgctggc aacagccctg caaggggctt ggctctgcca cagggcagga 3540
gaggaagttg tagcgcctag cgagagttcc agccccagac gcccacctgt gcctcagggc 3600
accgcctgcc gagcagagaa ggcacagcag ccgtcagagt ccatgagagg tgaaaccaca 3660
cagcagggat gtccaatatc agaactatta atatcaataa aagtataacc ttcccaggtc 3720
tatgcccaag agaattgaaa acatccatcc acacaatacc tgtgctcccg cgttcatagc 3780
agcattactc aaaagtcaaa cggtagcaac aacccaaatg tccatccaca gatgaattaa 3840
gacatgaagt gtgttctgtc catacaatgg aatattattt ggccataaaa aggaaggaaa 3900
ttctgacgca tgccacagcc tgagtgaatc ctacaaatat tacgctaagt gaaagaagcc 3960
aatcacgagt ttatgtgaaa tgtccagaat aggcaaatct gtgtatcaga gacaaagcac 4020

attggtggtt gccaggtact ggaggaagag agaagaggca tgacagctaa cagggacggg 4080
ctttctttgg aagatgatga aattgtggaa tgatggt 4117

<210> 30
<211> 2340
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5519834CB1

<400> 30
ctagtctcca tctcgtttac tcgggatggg acaagacgac caaaaaagag aggattcctc 60
atcaggttta ttactcgttg ccattcttgc aatgattacc gcccccgttg ctgtggttgg 120
gggaggcgag caaagggcag cggctgcgag cgcccgcccc cgccccaccc tcccagcccc 180
ggacagcgca ggctgcggct tttcgtcctc cactgagtcc tgccggtggc ccgagcccgg 240
tggcctcccg gcgaccctcg gcgcgaggcg acatggcagg cggccacagc ctcctgctgg 300
agaacgcgca gcaagtggtg ctggtgtgcg cccgcggcga gcgcttcctg gcgcgggatg 360
cgctgcgcag cctggcggtg ctggaaggcg ccagcctggt ggtgggcaaa gatggattta 420
taaaagctat gggtcctgct gatgttattc aaagacagtt ttctggagaa acttttgaag 480
aaataattga ctgctctggg aaatgtattc taccaggttt ggtggatgca cacacacatc 540
cagtatgggc tggtgaaaga gttcacgaat ttgcaatgaa gttggcagga gccacctaca 600
tggaaattca ccaggccgga ggagggatcc actttaccgt ggagcgcacg cgccaagcca 660
cagaggagga gctgttccgc tccttgcagc aacggctcca gtgcatgatg agggctggca 720
ccacgctggt ggagtgcaag agtggatatg gcctcgacct ggagaccgag ctcaagatgc 780
tgcgcgtgat tgagcgcgcc cggcgggagc tggacatcgg catctcggct acctactgcg 840
gggctcattc agtgcctaaa ggaaaaactg ctactgaagc tgctgatgac atcatcaata 900
accacctccc aaagctgaag gaacttggca gaaatgggga aatacacgtg gacaatatag 960
acgtattttg tgagaaaggt gtctttgatc tcgattccac cagaaggatt cttcaacgtg 1020
gaaaagatat agggttacag attaacttcc atggggatga actccacccg atgaaggctg 1080
ctgagcttgg ggctgaactg ggagcgcagg caatcagcca cctggaagaa gtgagtgatg 1140
aaggcatcgt tgccatggca acggccaggt gctctgccat ccttctgccc accacagcct 1200
acatgctgag actgaaacaa cctcgagcca ggaagatgtt agatgaagga gtaatagttg 1260
ctctgggaag tgatttcaac cccaatgcat attgcttttc aatgccaatg gtcatgcatc 1320
tggcctgtgt aaacatgaga atgtccatgc ctgaggcctt ggccgctgcc accatcaatg 1380
cagcttatgc actgggaaag tctcacacac acggatcgtt ggaagttggc aaacagggag 1440
atctcattat catcaattca tcccgatggg agcatttgat ttaccagttc ggaggccatc 1500
atgaattaat tgaatatgtt atagctaaag gaaaactcat ctataaaaca tgatagattt 1560
gaaaagagaa gacttttga ctatatgaaa taagtcaata tagttatatt aaaagttaaa 1620
acaccttaat atttacaaga attatatcac ttaaacctaa atgtacttca atgtcttttt 1680
aagtcactca aaaaacccaa gggatagatt tattttcatt taacacatgc atttgacata 1740
taaacaggta aacctattgt gattaaaatc acaaaacatc caattagttc acaaatattg 1800
gttacaaata ttctgtagat taatatggtg gggtatcaca aaaatgcctt tgtggggaaa 1860
agtaggcttg gcttaaaatt tccattttgt gtctgtattt cacatctcag tttttaaact 1920
atatttagtg aacattgagg gatcgaaaga aatctaagtg atacgcccca atgaagctaa 1980
aatatagcct tctgttaagc aaatagtatt tcctttcccc aagtagttca ttttctagat 2040
gcttgtcaaa tgaattaatg tcctctgatg aagagtgtcc ttccgtttct aaggtcttct 2100
caatctcagc aatagagctt cccagcagcg ttcaagacac atcatttata cacaggcaca 2160
ggggccttcc tgaaatgggt gcatttttac caactacaat catgtaattt ttttggaaat 2220
tttttaaaat tttcgattct ttacattaca attgggtgaa acacatttta cagctctcaa 2280
taaatgtttg ctgtcgctct taaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2340

<210> 31
<211> 2634
<212> DNA

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2215017CB1

<400> 31
atggatagcc ttaagcaaga aaataaaaat gatagagcca aaaagaaaga ccaatttaaa 60
aagggcagaa ttgggaacaa ggtccaaaca atcaaaaaga ataagaggtg taaaccttca 120
agtgctggaa gaaagaagcc aggaatgtat actgacagca ttaataagga cacaaagcct 180
ccccacatca tcttcatcct cacggacgac caaggctacc acgacgtggg ctaccatggt 240
tcagatatcg agacccctac gctggacagg ctggcggcca agggggtcaa gttggagaat 300
tattacatcc agcccatctg cacgccttcg cggagccagc tcctcactgg caggtaccag 360
atccacacag gactccagca ttccatcatc cgcccacagc agcccaactg cctgcccctg 420
gaccaggtga cactgccaca gaagctgcag gaggcaggtt attccaccca tatggtgggc 480
aagtggcacc tgggcttcta ccggaaggag tgtctgccca cccgtcgggg cttcgacacc 540
ttcctgggct cgctcacggg caatgtggac tattacacct atgacaactg tgatggccca 600
ggcgtgtgcg gcttcgacct gcacgagggt gagaatgtgg cctgggggct cagcggccag 660
tactccacta tgctttacgc ccagcgcgcc agccatatcc tggccagcca cagccctcag 720
cgtcccctct tcctctatgt ggccttccag gcagtacaca cacccctgca gtccctcgt 780
gagtacctgt accgctaccg caccatgggc aatgtggccc ggcggaagta cgcggccatg 840
gtgacctgca tggatgaggc tgtgcgcaac atcacctggg ccctcaagcg ctacggtttc 900
tacaacaaca gtgtcatcat cttctccagt gacaatggtg gccagacttt ctcgggggc 960
agcaactggc cgctccgagg acgcaagggc acttattggg aaggtggcgt gcggggccta 1020
ggctttgtcc acagtcccct gctcaagcga aagcaacgga caagccgggc actgatgcac 1080
atcactgact ggtacccgac cctggtgggt ctggcaggtg gtaccacctc agcagccgat 1140
gggctagatg gctacgacgt gtggccggcc atcagcgagg gccgggcctc accacgcacg 1200
gagatcctgc acaacattga cccactctac aaccatgccc agcatggctc cctggagggc 1260
ggctttggca tctggaacac cgccgtgcag gctgccatcc gcgtgggtga gtggaagctg 1320
ctgacaggag accccggcta tggcgattgg atcccaccgc agacactggc caccttcccg 1380
ggtagctggt ggaacctgga acgaatggcc agtgtccgcc aggccgtgtg gctcttcaac 1440
atcagtgctg acccttatga acgggaggac ctggctggcc agcggcctga tgtggtccgc 1500
accctgctgg ctcgcctggc cgaatataac cgcacagcca tcccggtacg ctacccagct 1560
gagaaccccc gggctcatcc tgactttaat gggggtgctt gggggccctg ggccagtgat 1620
gaggaagagg aggaagagga agggagggct cgaagcttct cccggggtcg tcgcaagaaa 1680
aaatgcaaga tttgcaagct tcgatccttt ttccgtaaac tcaacaccag gctaatgtcc 1740
caacggatct gatggtgggg agggagaaaa ctgtccttta gaggatcttc cccactccgg 1800
cttggccctg ctgtttctca gggagaagcc tgtcacatct ccatctacag ggagttggag 1860
ggtgtagagt cccttggttg aacagggtag ggagcctgga taggagtggg tgggaataaa 1920
ccagactggg atgcctgtgt ctcagtcctg cctcctcacg gacttgctct gtgacctcag 1980
gtgacccaca tgagctttta gcctcagttt cctcatctgt aaaatgagct ctaatgactt 2040
tgtgactctt tggtgtggcc ctggagcctg gggccacggt ggagttcctg gccggccttg 2100
ccacttgaca actcctttaa ggcttccccc ttaacacggg atccctgtgg tggtgtttgg 2160
gagttgcctg gaggcaactc caagcctggc ccccagctga agcatggcaa tctggctgct 2220
ctctacaggg acccccaagc gctgtgggtg gagggcaggg gtcgggggggg ttgaccttct 2280
tgggtcttca catggcctag gccagtcctc cggtcagact ggtgtcaggc accgtggtgc 2340
aaaattcctc ttctggcccc tccagtaccc cagagaaact ggctgggcca ttaactgctg 2400
cagcaccaag ggtggtagaa agagctgtga agagccccca aaccagtacc aggacacctg 2460
ggttctcctg tgacctgggg cacagttctt gccctctagg ccttgatttc cccacctgca 2520
agtggggatg ccagccctgg ctctgcctcc ttcatgaggc tctggaagac tggccaaggt 2580
tgtggaggag cttgtgaact tgattaaagt gtcgtaacat ggaaaaaaaa aaaa 2634

<210> 32
<211> 1188
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7484731CB1

<400> 32
atggtctacc ctggctcctc actcatcaga ggaaatgacg agattccaag gtcccctccg 60
gtattgcaag ctggcctgag ccatagggtc ctgtcccagc atggcatcaa atgcctggag 120
ctcatcctgc aggagcagcc atggcggagc ccccagacca aggcccagtc cccttcacct 180
gtctccatca taagcgtcac cagcctgtct gacaaactgc ccagcaaagt tattgacatt 240
ggcctggccg cctggggatg caaggccctg aacattgtag agaatgagat gccaggcctg 300
atgcacatgt gggagctgta ctcggcctcc aagccactag agggcaccca cagtgccagc 360
tgcctgcaca tgaccatgga gatggccatc cacattgggt ccctcatcac cctgggtgcc 420
ccagcagcaa catcttcctc accctggacc atgtgttggc tgccattgcc aaggctggca 480
ttccagttta cacctggaag ggcaaaagaa cgcagagtac ccatggtgca tcgagcagtc 540
actgtaattc agggatgggc tctcaacatt attctggagg atgggggtga ccttgccaac 600
ctcttccaca ccaagatgat ggtcaatggg atcctgaagg tgcctgccat caatgtcaat 660
gactccctca ccaagagtga gttcaacaag ctctatggct gctgggagtc cctcatagat 720
ggcatcaagt gggccacagt ggtgatgatt gccggcaagg tagcgatggt agcaggctat 780
ggcaatgtgg gcaagggctg tgcccaggcc ctgtggggtt tcggggccca cgtcatcatc 840
accaagatcg accccatcaa tgcactgcag gctgccatgg agggctatga ggtgaccacc 900
atggatgaag cctgtcagga gggcaacatc tttatcacca ccacagcctg tgtcaacatc 960
atccttggcc ggcactttga acagatgaag gatgatgcca ttgtatgtaa cactggacac 1020
tttgaggtgg agatcaatgt caagtggctc aagaagtacc ccattgaggt tcatgtccta 1080
cccaagaagc tggatgaggc agtggctgaa gcccacctgg gcaagctgat catgaaattg 1140
accaagctga ctgagaaaca ggcccagtac ctggggcatc tcccctga 1188

<210> 33
<211> 1670
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3927361CB1

<400> 33
atgcagcagc agcagcagca aaaacagaaa cagccaggaa gatcacctga attttcaggc 60
gcacttgccc tgcaagccct tctggaaggc ggcacgtcca ggcggtccag cgcatcgagc 120
cgcgccaggc gacccgagat gcccccgaat ctgtccagct cgccctgcag atcgcacgct 180
ccaactggcg gattccgacg taaccctgtg cgcccgaggc ccagcggtaa cccgccgaag 240
cgggggcggt accttgtgac catgtcttcc aagaagaata gaaagcggtt gaaccaaagc 300
gcggaaaatg gttcgtcctt gccctctgct gcttcctctt gtgcggaggc acgggctcct 360
tctgctggat cagacttcgc ggcaacctcc gggactctga cggtgaccaa cttattagaa 420
aagggtaaag aattccgggt gtatacagcc tggcctatgg caggatttcc tggaggcaag 480
gtcggcctga gtgaaatggc acagaaaaat gtgggtgtga ggcctggtga tgccatccag 540
gtccagcctc ttgtgggtgc tgtgctacag gctgaggaaa tggatgtggc actgagtgac 600
aaagatatgg aaattaatga agaagaactg actggttgta tcctgagaaa actagatggc 660
aagattgttt taccaggcaa ctttctgtat tgtacattct atggacgacc gtacaagctg 720
caagtattgc gagtgaaagg ggcagatggc atgatattgg gagggcctca gagtgactct 780
gacactgatg cccaaagaat ggcctttgaa cagtccagca tggaaaccag tagcctggag 840
ttatccttac agctaagcca gttagatctg gaggataccc agatcccaac atcaagaagt 900
actccttata aaccaattga tgacagaatt acaaataaag ccagtgatgt ttgctggatg 960
tacacagagc cctggagatg gcagtggacc aatcaaaaag gattgttact ttatggtcct 1020
ccatgtactg gaaaaacaat gatcgccagg gctgttgcta atgaatttgg agcctatgtt 1080
tctgtaatta atggtcctga aattataagc aaacacccat caattatttt tattgatgag 1140
ctggatgcac tttgtccgaa aagagagggg gcccagaatg aagtggaaaa aagagttgtg 1200

gcttcactct taacactgat ggatggcatt ggttcagaag taagtgaagg acaagtgttg 1260
gttcttgggg gcacaaatcg gcctcatgcc ttggatgctg ctctccgaag acctgggcga 1320
tttgataaag agattgagat tggagttccc aatgctcagg accggctaga tattctccag 1380
aaactgcttc gagtacccca tttgctcact gaggctgagc tgctgcagct ggcaaatagt 1440
gctcatggat acgttggagc agacttgaaa gtcttgtgta atgaagcagg tgagtgtggt 1500
ttgctatggt gagtctctat tgatgcactt atctccagtt tacttacata caaataattt 1560
atattttaca gatttcttaa tggaagtagc tttgtttcta attataaaat gtgtaatttt 1620
tatttgaaaa aatttttatt ggaaaaacct agatgattca gaagactata 1670

<210> 34
<211> 1070
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6542758CB1

<400> 34
ctgtgagcct cagtttcctc atgaatgtaa tagatatgag aacactatga ttttgcagtc 60
ctggtttcta cgtggatcta atggtgtaat ggatatgaat gtacattgca agctgttaag 120
tcatttctgc taaatacaga atgacccact gtggtcatct tggaagcctc ggccatccct 180
gctggctgtg ctccattctt gtgcctcgcc attggaacgc tctagtgagc cggaatgaag 240
ttcaggccca tggctgtgat gtcacagaac atgtgaagtc agaggtccta tggaaggtga 300
ggggagaaaa tgcccctgga aagggttaag ggccaggaca ggaatggggc aggaggtgca 360
cggatcctgc tgggcactgg gagcaggggg cggccaaagg cagtgggtgg gcaggtccat 420
gcctcccctg gccccccagc tctgcagggc agtgttcctg gttcctatct tgctgctgct 480
gcaggtgaag cctctgaacg ggagcccagg ccccaaagat gggagccaga cagagaaaac 540
gccctctgca gaccagaatc aagaacagtt cgaagagcac tttgtggcct cctcagtggg 600
tgagatgtgg caggtggtgg acatggccca gcaggaagaa gaccagtcgt ccaagacggc 660
agctgttcac aagcactctt tccacctcag cttctgcttt agtctggcca gtgtcatggt 720
tttctcagga gggccattga ggcggacatt cccaaatatc caactctgct tcatgctcac 780
tcactgaccc tccctccctc ctgggctcca ggtcacaact cccaaaggag atgcaggcat 840
ggctctctgc ctctgatcac catcactgta tctcaaggtt cagcagcaga gataccagtt 900
gccatcagtg ctaactgact gcctctccag gttcggagtt tcatctccca gggccagaga 960
cagcagaccc acatccttct ctcccacacc tctcctggtt ttgttcagga cagcagatta 1020
gaggcaggag gcaatgacaa taaaataacg ataaaatcct gaaaaaaaaa 1070

<210> 35
<211> 2000
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3188878CB1

<400> 35
gcagttcctc ccggggtcgg aggccgattc gccgtgtggc gggttcgagt cccgcctcct 60
gactctggcc tctagtccct gagtcccggg cgggctgcat tcgtcgggga aacctctcct 120
cgaccagggg cacctctact cgaccagggg cgacggcgta ctttgggctt catcatggag 180
gactacctgc agggttgtcg agctgctctg caggagtccc gacctctaca tgttgtgctg 240
ggaaatgaag cctgtgattt ggactccaca gtgtctgctc ttgccctggc tttttaccta 300
gcaaagacaa ctgaggctga ggaagtcttt gtgccagttt taaatataaa acgttctgaa 360
ctacctctgc gaggtgacat tgtcttcttt cttcagaagg ttcatattcc agagagtatc 420
ttgatttttc gggatgagat tgacctccat gcattatacc aggctggcca actcaccctc 480

atccttgtcg accatcatat cttatccaaa agtgacacag ccctagagga ggcagtagca 540
gaggtgctag accatcgacc catcgagccg aaacactgcc ctccctgcca tgtttcagtt 600
gagctggtgg ggtcctgtgc taccctggtg accgagagaa tcctgcaggg ggcaccagag 660
atcttggaca ggcaaactgc agcccttctg catggaacca tcatcctgga ctgtgtcaac 720
atggacctta aaattggaaa ggcaacccca aaggacagca aatatgtgga gaaactagag 780
gccctttcc cagacctacc caagagaaat gatatatttg attccctaca aaaggcaaag 840
tttgatgtat caggactgac cactgagcag atgctgagaa aagaccagaa gactatctat 900
agacaaggcg tcaaggtggc cattagtgca atatatatgg atttggaggc ctttctgcag 960
aggtctaacc tccttgcaga tctccatgct ttctgccagg ctcacagcta tgatgtcctg 1020
gttgccatga ctatcttttt caacactcac aatgagccag tgcggcagtt ggctattttc 1080
tgtccccatg tggcactcca aacaacgatc tgtgaagtcc tggaacgctc ccactctcca 1140
cccctgaagc tgacccctgc ctcaagtacc caccctaacc tccatgccta tcttcaaggc 1200
aacacccagg tctctcgaaa gaaacttctg cccctgctcc aggaagccct gtcagcatat 1260
tttgactcca tgaagatccc ttcaggacag cctgagacag cagatgtgtc cagggagcaa 1320
gtggacaagg aattggacag ggcaagtaac tccctgattt ctggcctgag tcaagatgag 1380
gaggaccctc cgctgccccc gacgcccatg aacagcttgg tggatgagtg ccctctagat 1440
caggggctgc ctaaactctc tgctgaggcc gtcttcgaga agtgcagtca gatctcactg 1500
tcacagtcta ccacagcctc cctgtccaag aagtgactgt tgagaggcga ggaggtagtg 1560
ggtgaggcta cctgactcac ttcaaatgca tgttttgaga tgtttggaga ttcagcaatt 1620
ctgtcttcat tgctccagga tctggtatac tgttctcata aaactgagag gagaaaaaaa 1680
gtgaaagaaa gcagctgctt taagaatggt tttccacctt ttccccctaa tctctaccaa 1740
tcagacacat tttattattt aaatctgcac ctctctctat tttatttgcc aggggcacga 1800
tgtgacatat ctgcagtccc agcacagtgg gacaaaaaga atttagaccc caaaagtgtc 1860
ctcggcatgg atcttgaaca gaaccagtat ctgtcatgga actgaacatt catcgatggt 1920
ctccatgtat tcatttattc acttgttcat tcaagtattt attgaatacc tgcctcaagc 1980
tagagagaaa agagagtgcg 2000

<210> 36
<211> 2559
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7500488CB1

<220>
<221> unsure
<222> 465-512
<223> a, t, c, g, or other

<400> 36
agcatgtgtg caaagtctat gcatctctga ccttgggtgc tgaggatcag gaaccgacct 60
actgcaacat gggccacctc agtagccacc tccccggcag ggccctgagg agcccacgga 120
atacagcacc atcagcaggc cttagcctgc actccaggct ccttcttgga ccccaggctg 180
tgagcacact cctgcctcat cgaccgtctg ccccctgctc ccctcatcag gaccaacccg 240
gggactggtg cctctgcctg atcagccagc attgccccta gctctgggtt gggcttgggg 300
ccaagtctca gggggcttct aggagttggg gttttctaaa cgtcccctcc tctcctacat 360
agttgaggag ggggctaggg atatgctctg gggctttcat gggaatgatg aagatgataa 420
tgagaaaaat gttatcatta ttatcatgaa gtaccattat cagannnnnn nnnnnnnnnn 480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nngcagttcc tcccggggtc ggaggccgat 540
tcgccgtgtg gcgggttcga gtcccgcctc ctgactctgg cctctagtcc ctgagtcccg 600
ggcgggctgc attcgtcggg gaaacctctc ctcgaccagg ggcacctcta ctcgaccagg 660
ggcgacggcg tactttgggc ttcatcatgg aggactacct gcagggttgt cgagctgctc 720
tgcaggagtc ccgacctcta catgttgtgc tgggaaatga agcctgtgat ttggactcca 780
cagtgtctgc tcttgccctg gctttttacc tagcaaagac aactgaggct gaggaagtct 840

```
ttgtgccagt tttaaatata aaacgttctg aactacctct gcgaggtgac attgtcttct 900
ttcttcagaa ggttcatatt ccagagagta tcttgatttt tcgggatgag attgacctcc 960
atgcattata ccaggctggc caactcaccc tcatccttgt cgaccatcat atcttatcca 1020
aaagtgacac agccctagag gaggcagtag cagaggtgct agaccatcga cccatcgagc 1080
cgaaacactg ccctccctgc catgtttcag ttgagctggt ggggtcctgt gctaccctgg 1140
tgaccgagag aatcctgcag ggggcaccag agatcttgga caggcaaact gcagcccttc 1200
tgcatggaac catcatcctg gactgtgtca acatggacct taaaattgga aaggcaaccc 1260
caaaggacag caaatatgtg gagaaactag aggccctttt cccagaccta cccaagagaa 1320
atgatatatt tgattcccta caaaaggcaa agtttgatgt atcaggactg accactgagc 1380
agatgctgag aaaagaccag aagactatct atagacaagg cgtcaaggtg gccattagtg 1440
caatatatat ggatttggag atctgtgaag tcctggaacg ctcccactct ccacccctga 1500
agctgacccc tgcctcaagt acccacccta acctccatgc ctatcttcaa ggcaacaccc 1560
aggtctctcg aaagaaactt ctgcccctgc tccaggaagc cctgtcagca tattttgact 1620
ccatgaagat cccttcagga cagcctgaga cagcagatgt gtccagggag caagtggaca 1680
aggaattgga cagggcaagt aactccctga tttctggact gagtcaagat gaggaggacc 1740
ctccgctgcc cccgacgccc atgaacagct tggtggatga gtgccctcta gatcaggggc 1800
tgcctaaact ctctgctgag gccgtcttcg agaagtgcag tcagatctca ctgtcacagt 1860
ctaccacagc ctccctgtcc aagaagtgac tgttgagagg cgaggaggta gtgggtgagg 1920
ctacctgact cacttcaaat gcatgttttg agatgtttgg agattcagca attctgtctt 1980
cattgctcca ggatctggta tactgttctc ataaaactga gaggagaaaa aaagtgaaag 2040
aaagcagctg ctttaagaat ggttttccac cttttccccc taatctctac caatcagaca 2100
cattttatta tttaaatctg cacctctctc tattttattt gccaggggca cgatgtgaca 2160
tatctgcagt cccagcacag tgggacaaaa agaatttaga ccccaaaagt gtcctcggca 2220
tggatcttga acagaaccag tatctgtcat ggaactgaac attcatcgat ggtctccatg 2280
tattcattta ttcacttgtt cattcaagta tttattgaat acctgcctca agctagagag 2340
aaaagagagt gcgctttgga aatttattcc agttttcagc ctacagcaga ttataagccc 2400
gggagctttt ttttggcgcc ccatgtgttg gggtcgttcc aaaagcggat cactctacca 2460
ctatggggtc cccactcttg gggcaatagc gagtttttc tcaaaacgcg gttttttccc 2520
tccccccccc ccttttttt aaaccccgt ttttcttca                        2559
```